G-GCSN: Global Graph Convolution Shrinkage Network for Emotion Perception from Gait

Yuan Zhuang[1], Lanfen Lin[1], Ruofeng Tong[1], Jiaqing Liu[2], Yutaro Iwamoto[2], and Yen-Wei Chen[3,2,1]

[1] College of Computer Science and Technology, Zhejiang University, 310000, China
zhuangyuan97@zju.edu.cn, llf@zju.edu.cn, trf@zju.edu.cn
[2] College of Information Science and Engineering, Ritsumeikan University, 525-8577, Japan gr0302kv@ed.ritsumei.ac.jp, yiwamoto@fc.ritsumei.ac.jp, chen@is.ritsumei.ac.jp
[3] Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, 310000, China

Abstract. Recently, emotion recognition through gait, which is more difficult to imitate than other biological characteristics, has aroused extensive attention. Although some deep-learning studies have been conducted in this field, there are still two challenges. First, it is hard to extract the representational features of the gait from video effectively. Second, the input of body joints sequences has noise introduced during dataset collection and feature production. In this work, we propose a global link, which extends the existing skeleton graph (the natural link) to capture the overall state of gait based on spatial-temporal convolution. In addition, we use soft thresholding to reduce noise. The thresholds are learned automatically by a block called shrinkage block. Combined with the global link and shrinkage block, we further propose the global graph convolution shrinkage network (G-GCSN) to capture the emotion-related features. We validate the effectiveness of the proposed method on a public dataset (i.e., Emotion-Gait dataset). The proposed G-GCSN achieves improvements compared with state-of-the-art methods.

1 Introduction

Emotion recognition is of great value in an intelligent system to insight into human psychology. Current most studies of emotional recognition use facial expressions [1][2], speech (including words and sounds) [3][4], and physiological signals (e.g., EEG, heart rate, breathing, body temperature, etc.) [5][6][7]. Through the analysis and processing of these data, emotions are divided into emotional categories such as sadness, happiness, anger, etc.

Recently, some studies have shown that emotional expression is also reflected in body language. Body language includes body posture and body movement, in which gait is the most basic body language. Gait refers to the modalities displayed during human walking, consisting of a series of body joints in time sequence. Gait can reflect emotional information, and individuals in different

emotional states may have differences in some gait characteristics, such as arm swing, speed, angle, length of steps, etc. [8][9][10].

Compared with traditional biological characteristics for emotion detection, such as facial expressions, speech, and physiological signals, gait is observable remotely, more difficult to imitate, and does not require a high degree of participation of the subject. In the real environment, it can be easily collected by a non-contact method,which is convenient to be applied to all kinds of scenes in real life.

So, this paper focuses on the relationship between gait and emotion. Some published studies applied models for the action recognition task, such as STEP [11]. In STEP, they applied spatial-temporal graph convolution and used the physical structure of the human skeleton as the spatial graph, which only captures the local dependencies without considering the non-local correlation of joints. Besides, the published studies ignored the noise of the input. The gait is inputted as a point cloud composed of a series of body joints in time sequence extracted from the video. External environment and extraction algorithms may introduce noise that can impact on our tasks.

In this work, we mainly solve the above two problems. First, we propose a global link, a new joint connection method of constructing the skeleton graph which is suitable for emotion recognition to perceive emotions as a whole. Second, we propose a graph convolution shrinkage (GCS) block, a new module that combines spatial graph convolution and temporal graph convolution with shrinkage block to reduce noise impacts. Based on the global links and GCS, we propose a global graph convolution shrinkage network (G-GCSN), stacked by multiple global graph convolution shrinkage blocks.

To verify the effectiveness of the proposed G-GCSN, we carry out experiments on Emotion-Gait[11], which is the same dataset as used in STEP[11]. The main contributions in this work are summarized as followed:

(1) We propose a global link that constructs a new skeleton graph to complement the original natural links in physical structures to capture the overall state of gait based on spatial-temporal convolution.

(2) We propose a graph convolution shrinkage block (GCS) to reduce the influence of the noise.

(3) We stack global graph convolution shrinkage blocks to give an outperformance network (G-GCSN) to extract emotion-related features effectively.

2 Related Work

Walking is the most basic activities of daily life and the gait displayed during walking varies from person to person. Thence, gait information can be used as a unique biological feature for identity detection [12][13] applied to security and other scenes. Moreover, some researches [14][15][16] used gait information to detect abnormal movements and gait patterns, which can be used as the basis for early diagnosis of mental diseases (such as cerebral palsy, Parkinson's

disease, Rett syndrome, etc.) in clinical. And gait is also used in action recognition [17][18], mainly applied in video surveillance, intelligent human-computer interaction, virtual reality, etc.

Recently, the gait has been used for emotion recognition. The published studies on gait-based emotion recognition mostly adopt machine learning methods. Li et al. [19] extracted the time and frequency domain features of the six nodes on the arms and legs, and used SVM and LDA for classification; Quiroz et al. [20] adopted sliding windows to extract feature vectors, and used Random Forest for classification. Ahmed et al. [21] extracted geometric and kinematic features from an LMA framework and adopted KNN, SVM, LDA and other classifiers combined with Score-level Fusion and Rank-level Fusion for feature fusion. Zhang et al. [22] extracted 3D acceleration data from the right wrist and ankle and used Decision tree, SVM, Random Forest, Random Tree as classifiers. Venture et al. [23] extracted the joint angles and classified them based on similarity. Karg et al. [24] used PCA-based classifiers on some extracted features such as stride length, velocity and so on. Daoudi et al. [25] represented body joints movement as symmetric positive definite matrices and performed the nearest neighbor classification. Wang et al. [26]fed low-level 3D postural features and high-level kinematic and geometrical features to a Random Forest classifier. Crenn et al. [27] computed geometric features, motion features, and Fourier features of body joints movement and used the support vector machine with a radial basis function kernel as the classifier.

There are also a few studies adopting deep learning methods. Randhavane et al. [28]. used RNN-based LSTM to extract deep features and combined hand-extracted emotional features. Random Forest Classifier is used to classification. Bhattacharya et al.[11]. used the features extracted by the ST-GCN network and combined with hand-extracted emotional characteristics for classification. In this work, we only use deep learning method which based on GCN (ST-GCN) [29] for feature extraction, without manually extracted features.

3 Background

ST-GCN ST-GCN [29], the spatial-temporal graph convolution network, is composed of multiple spatial-temporal graph convolution blocks. Each block consists of two parts: spatial graph convolution and temporal graph convolution, which are used to obtain the features in space and time dimensions respectively. The crucial part of spatiotemporal graph convolution is constructing the spatial-temporal graph (Fig 1(a)). The spatial graph is the natural connection of body joints. According to the spatial conguration partition strategy in ST-GCN, it can be divided into three subgraphs (show in Fig 1(b)): 1) the root node itself 2) centripetal group 3) centrifugal group. The neighbor relation of each subgraph is represented by an adjacency matrix. And the time graph is structed by the connection of the same node between adjacent frames. Let $\mathbf{X_{in}} \in \mathbb{R}^{\mathbf{c_{in}*t_{in}*v}}$ be the input features of all joints in all frames and $\mathbf{X_{out}} \in \mathbb{R}^{\mathbf{c_{out}*t_{out}*v}}$ be the output, where c_{in}, c_{out} is the dimension of input joints features and output joints

features, t_{in}, t_{out} is the number of input frames and output frames, and v is the number of joints. The spatial graph convolution is formulated as:

$$\boldsymbol{X_{out}} = \sum_{\boldsymbol{k}=0}^{\mathrm{K}} (\boldsymbol{M_k} \odot \boldsymbol{A_k}) \boldsymbol{X_{in}} \boldsymbol{W_k} \tag{1}$$

where K is the number of subgraphs. $\mathbf{A_k} \in \mathbb{R}^{\mathbf{v*v}}$ is the adjacency matrix of the subgraph k. For the adjacency matrix $\mathbf{A_k}$, it has a weight matrix $\mathbf{M_k} \in \mathbb{R}^{\mathbf{v*v}}$. $\odot$ donates element-wise product between two matrixes.$\mathbf{W_k}$ is a weight function. $\mathbf{M_k}$ and $\mathbf{W_k}$ are trainable.

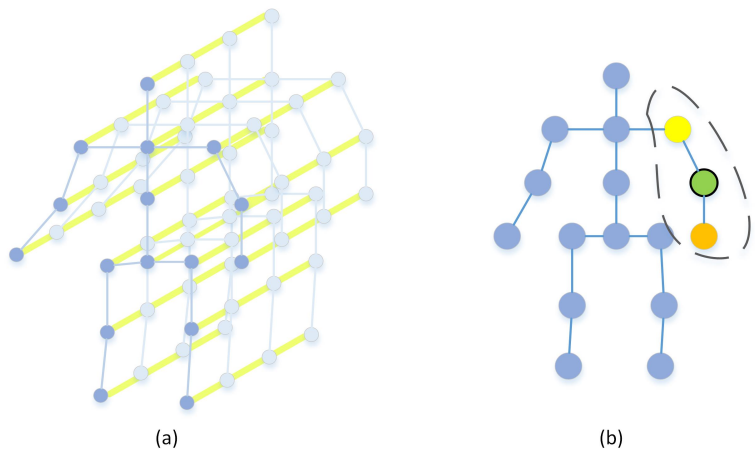

Fig. 1. (a) The spatial-temporal graph used in ST-GCN. (b) Illustration of the partition strategy and different color nodes are in different subgraphs. (The green one is the root node, the yellow one and the orange one are in the centripetal group and centrifugal group respectively.)

4 Proposed Method

Our work is to perceive the emotion of gait based on spatial-temporal graph convolution. The final architecture is shown in Fig 2. We feed the 3D coordinates of body joints extracted from the walking video into the network in a time sequence. And the output is an emotion classification result of the gait. The proposed global graph convolution shrinkage network (G-GCSN) consists of three GCS blocks,followed by a global average pooling (GAP) layer, a fully connected (FC) layer, and a softmax function. The base block is the graph convolution shrinkage block GCS which comprises three components: (1) Spatial graph convolution. The original spatial graph is constructed by the natural link.

The natural link only captures the local dependencies, which may miss the non-local correlation of joints. We proposed a global link to capture the overall state of gait by the non-local dependencies. The final spatial graph is constructed by our proposed global link and the natural link. (2) Temporal graph convolution. Its graph is formed by the connection of the same node between the adjacent frames. We use it to extracted features in the time dimension. (3) Shrinkage block. The input data may be mixed with noise during the collection process, and the noise may be further amplified in the network. We introduce it to reduce noise after spatial-temporal convolution. In sections 4.1 and 4.2, we will introduce the global link and shrinkage block in detail.

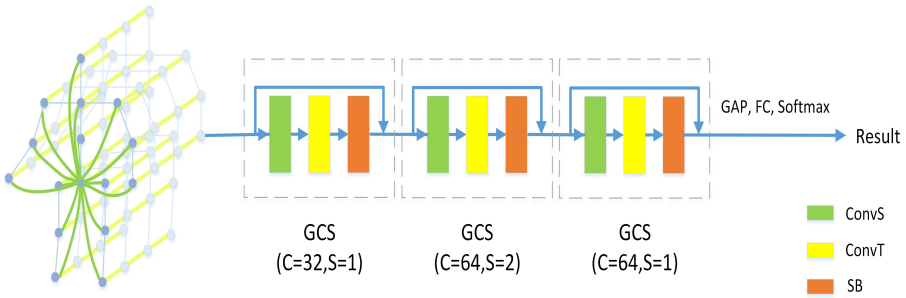

Fig. 2. Architecture Overview. GCS, ConvS, ConvT and SB denotes graph convolution shrinkage block, spatial graph convolution, temporal graph convolution, and shrinkage block. C represents the number of output channels for GCS. S represents the stride of temporal graph convolution. Each spatial graph convolution and temporal graph convolution is followed by a batch normalization layer and a ReLU layer which are not shown in the figure. Moreover, a residual connection is added for each GCS block.

4.1 Global Graph Convolution

Spatiotemporal graph convolution (ST-GCN) has made good progress in many fields, such as action recognition [30][31], traffic forecasting [32][33], and person re-identification [34]. Recently, ST-GCN has also been applied to emotion recognition based on gait by Bhattacharya et al.[11].

Bhattacharya et al. proposed a STEP network which is stacked by three spatiotemporal graph convolution blocks. There is no change in the spatial and temporal graph compared with ST-GCN. The spatial graph is constructed by the physical structure of the human skeleton (the natural link, shown in Fig 3(a)). And the natural link only captures the local dependencies, which may miss the non-local correlation of joints. Besides, the difference between the different emotional gaits is not the action but the overall state. Lets consider the influence of emotion on gait in an ideal situation. When happy, the body posture is stretched overall, the steps are light, the action amplitude is slightly bigger; When sad,

the body posture collapses overall, the steps are heavy, the action amplitude is slightly smaller; When angry, the body posture overall expands, the steps are strong, the action amplitude is slightly bigger. When neutral, the body is a little more stretched than when sad, the action amplitude is smaller than when happy and angry. So, it is significant to perceive gait as a whole for emotion recognition.

To perceive the emotion from the overall state of gait, we propose a global link representing the global dependencies to capture emotion-related features. We want to choose a stable center node of the skeleton as the reference joint. In this way, the overall state of gait can be simply perceived by the dependencies between the center node and other nodes. As shown in Fig 3(b), we choose the spine joint as the reference joint and obtain the global link by connecting the joint points to the spine joint.

The natural link and the global link complement each other and cannot be replaced. We add them together to get the complete spatial graph. And the Equ 1 can be replaced by

$$\boldsymbol{X_{out}} = \sum_{k=0}^{K} \left(\boldsymbol{M_k} \odot \left(\boldsymbol{A_n} + \boldsymbol{A_g} \right)_k \right) \boldsymbol{X_{in} W_k} \tag{2}$$

where $\mathbf{A_n} \in \mathbb{R}^{\mathbf{v*v}}$ is the adjacency matrix of natural link and $\mathbf{A_g} \in \mathbb{R}^{\mathbf{v*v}}$ is the adjacency matrix of the global link. K is the number of subgraphs, which are divided according to the spatial conguration partition strategy in ST-GCN.

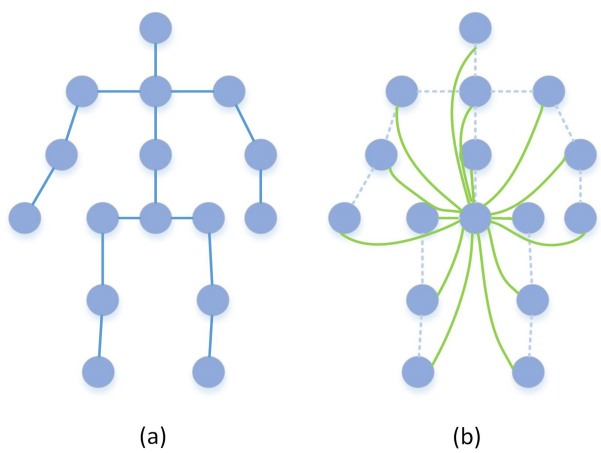

Fig. 3. The graph of body joints. (a)The natural link of body joints (b) The proposed global link of body joints (green lines)

4.2 Shrinkage Block

The gait is inputted as a point cloud composed of 3D coordinates of body joints in time sequence, which is extracted from videos. Noise may be introduced in the process of gait information collection due to the environment and the accuracy of the joint coordinate extraction algorithm. In the process of the spatial-temporal graph convolution, the noise is passed while transferring the features.

When operating convolution in the spatial dimension, each node passes the feature and noise to his neighbor nodes simultaneously through the adjacency matrix, which leads to the introduction of noise in each channel to some extent. At the same time, the convolution operation makes a fusion between channels. Then the noise of each channel will be fused into the new channel after convolution. The same is true in the temporal convolution. These noises have a negative impact on the experimental results.

Inspired by DRSN [35], we introduce a subnetwork called shrinkage block in the ST-GCN block after temporal convolution to reduce noise by soft thresholding. It is crucial to set the threshold. For any two samples, their noise is different. And there are also differences in noise from channel to channel. It is difficult to set the threshold for them all. We introduce a subnet to automatically learn thresholds for each channel of each sample from the sample features. And soft thresholding is carried out according to thresholds. As the network deepens, noise decreases layer by layer, reducing the interference and negative impacts on emotion recognition tasks. The function of soft thresholding can be expressed as follows:

$$y = \begin{cases} x - \tau & x > \tau \\ 0 & \tau \leq x \leq \tau \\ x + \tau & x < -\tau \end{cases} \tag{3}$$

where x is the input feature, y is the output feature, and τ is the threshold. τ not only needs to be positive, but also cannot be too large to prevent the output of soft thresholding being zero.

As shown in the Fig 4, the proposed method is similar to DRSN-CW. In the sub-network, first, perform GAP global average pooling on the absolute value of the feature map to obtain a one-dimensional vector $\mathbf{A}$. The one-dimensional vector passes through two FC layers to obtain a scale vector $\mathbf{W}$, and the elements in $\mathbf{W}$ is scaled to (0, 1) through the sigmoid function. Then, an element-wise product operation is performed on $\mathbf{A}$ and $\mathbf{W}$ to get the threshold vector $\mathbf{T}$. According to $\mathbf{T}$, we use soft thresholding to reduce noise on the features map. The threshold is expressed as follows:

$$\boldsymbol{T} = \boldsymbol{W} \odot \boldsymbol{A} \tag{4}$$

where $\odot$ represents the element-wise product operation. There are c elements in the vector , corresponding to the threshold values of c channels. Every element meets the threshold requirements within the interval of (0, max).

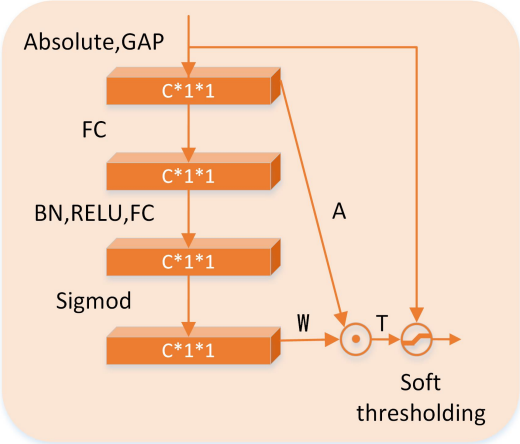

Fig. 4. The structure of shrinkage block (SB)

5 Experiments

In this section, we introduce the details of our experiments, including data sets, parameter settings, and ablation experiments.

5.1 Datasets and Implementation

Emotion-Gait (E-Gait) [11] is a public dataset, consisting of 2177 real gaits and 4000 generated gaits. Among the real gait samples, there are 342 collected from 90 participants and 1,835 from ELMD databases [36]. The 342 collected data came from the 90 participants who imaged four emotions (happy, sad, neutral, angry) while walking at a distance of 7 meters. These 2177 gaits are labeled as one of the four emotional categories by domain experts.

All experiments were performed in the PyTorch deep learning framework. For training G-GCSN, we randomly divide the data set into training set and test set according to the ratio of 9:1, and run ten times to get the average result. We use a batch size of 8 and train for 100 epochs using Adam optimizer with an initial learning rate of 0.001. The learning rate decreases to 1/10 of its current value after 50, 75, 88 epochs. We also use a momentum of 0.9 and weight-decay of .

We quantitatively evaluate the results of our experiments by classification accuracy given by the following formula:

$$Accuracy = (TH + TS + TN + TA)/TD \tag{5}$$

Where TH, TS, TN, and TA are the true positives number of four emotions, and TD is the number of total data.

5.2 Ablation Experiments

To validate the G-GCSN modules efficiency to capture emotion-related gait features, we build up the model by adding its individual components incrementally.

First, we evaluate the effectiveness of applying the global link to capture spatial information. In Table 1, (ST-)GCN indicates the network that uses the natural link as the spatial graph in ST-GCN block. G-GCN, the global graph convolution network, improves the spatial graph by adding the global link to the natural link. The G-GCN achieves better performance than the (ST-)GCN, which shows that the global link can effectively capture emotion-related features.

Second, to verify the effectiveness of the shrinkage block, we experiment on the global graph convolution shrinkage network (G-GCSN). G-GCSN is the model that combined shrinkage block into G-GCN. The experimental results in Table 1 show that the introduction of shrinkage block can effectively improve the performance of the network.

Table 1. The ablation experiment results.

Method	Accuracy
(ST-)GCN	78.24%
G-GCN	**81.10%**
G-GCSN	**81.50%**

Besides, we try some sizes of the temporal graph convolution kernel. As Table 2 shows, the other temporal kernel sizes do not perform better than 75 which we used for all of experiments in Table 1.

Table 2. The ablation experiment results.Model accuracy with various temporal kernel size k.

Method	Accuracy
G-GCSN	
with k=45	81.30%
with k=65	81.35%
with k=75	**81.50%**
with k=85	81.36%

5.3 Comparison with State-of-the-Art Methods

In the paper of STEP, they used some prior machine learning methods [23-27] for emotion recognition on Emotion-Gait. As Table 3 shows, the accuracy of all

Table 3. Comparison with state-of-the-art methods.

Method	Accuracy
Venture et al. [23]	30.83%
Karg et al. [24]	39.85%
Daoudi et al. [25]	42.52%
Wang et al. [26]	53.73%
Crenn et al. [27]	66.22%
ST-GCN [29]	65.62%
LSTM [28]	75.10%
Base-STEP[11]	78.24%
G-GCSN	**81.50%**

is less than 70%. And the accuracy on three action recognition from gaits methods [29][28][11] are 65.52%, 75.10% and 78.24%. By comparison, our G-GCSN achieved 81.50% on Emotion-Gait, which exceeds state-of-the-art performance.

6 Conclusions

In this paper, we propose a novel network called global graph convolution shrinkage network (G-GCSN) for human emotion recognition based on gait. The global link extracts the emotion-related gait features from the whole. And the shrinkage block can automatically set suitable thresholds through samples features and reduce the noise of the input features by soft thresholding. We validate G-GCSN in emotion recognition using a public dataset of Emotion-Gait and the results of experiments show that G-GCSN surpasses all previous state-of-the-art approaches.

However, we did not train our network on the synthetic gaits or combine with the affective features in Emotion-Gait dataset as Bhattacharya et al.[11] did in STEP and STEP+Aug. Our work is dedicated to replacing other data processing methods with network improvements to improve performance. In current work, the performance of our network on the real gaits is not as good as that of STEP and STEP+Aug. So the next step is to further improve the network structure. Given the difference between samples, we may design a more flexible spatial-temporal graph to replace the unified one. Besides, we will also verify the performance of our network on more datasets to prove its universal applicability.

Acknowledgements

This work was supported in part by Major Scientific Research Project of Zhejiang Lab under the Grant No.2018DG0ZX01, and in part by the Grant in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 20K21821.

References

1. Zhang, Y.D., Yang, Z.J., Lu, H.M., Zhou, X.X., Phillips, P., Liu, Q.M., Wang, S.H.: Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. IEEE Access **4** (2016) 8375–8385
2. Mohammadpour, M., Khaliliardali, H., Hashemi, S.M.R., AlyanNezhadi, M.M.: Facial emotion recognition using deep convolutional networks. In: 2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI), IEEE (2017) 0017–0021
3. Zhao, J., Mao, X., Chen, L.: Speech emotion recognition using deep 1d & 2d cnn lstm networks. Biomedical Signal Processing and Control **47** (2019) 312–323
4. Nicholson, J., Takahashi, K., Nakatsu, R.: Emotion recognition in speech using neural networks. Neural computing & applications **9** (2000) 290–296
5. Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. IEEE transactions on pattern analysis and machine intelligence **30** (2008) 2067–2083
6. Kim, K.H., Bang, S.W., Kim, S.R.: Emotion recognition system using short-term monitoring of physiological signals. Medical and biological engineering and computing **42** (2004) 419–427
7. Lin, Y.P., Wang, C.H., Jung, T.P., Wu, T.L., Jeng, S.K., Duann, J.R., Chen, J.H.: Eeg-based emotion recognition in music listening. IEEE Transactions on Biomedical Engineering **57** (2010) 1798–1806
8. Montepare, J.M., Goldstein, S.B., Clausen, A.: The identification of emotions from gait information. Journal of Nonverbal Behavior **11** (1987) 33–42
9. Halovic, S., Kroos, C.: Not all is noticed: Kinematic cues of emotion-specific gait. Human movement science **57** (2018) 478–488
10. Roether, C.L., Omlor, L., Christensen, A., Giese, M.A.: Critical features for the perception of emotion from gait. Journal of vision **9** (2009) 15–15
11. Bhattacharya, U., Mittal, T., Chandra, R., Randhavane, T., Bera, A., Manocha, D.: Step: Spatial temporal graph convolutional networks for emotion perception from gaits. In: AAAI. (2020) 1342–1350
12. Chao, H., He, Y., Zhang, J., Feng, J.: Gaitset: Regarding gait as a set for cross-view gait recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence. Volume 33. (2019) 8126–8133
13. Kale, A., Sundaresan, A., Rajagopalan, A., Cuntoor, N.P., Roy-Chowdhury, A.K., Kruger, V., Chellappa, R.: Identification of humans using gait. IEEE Transactions on image processing **13** (2004) 1163–1173
14. Gage, J.R.: Gait analysis. an essential tool in the treatment of cerebral palsy. Clinical orthopaedics and related research (1993) 126–134
15. Jellinger, K., Armstrong, D., Zoghbi, H., Percy, A.: Neuropathology of rett syndrome. Acta neuropathologica **76** (1988) 142–158
16. Jankovic, J.: Parkinsons disease: clinical features and diagnosis. Journal of neurology, neurosurgery & psychiatry **79** (2008) 368–376
17. Li, M., Chen, S., Chen, X., Zhang, Y., Wang, Y., Tian, Q.: Actional-structural graph convolutional networks for skeleton-based action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 3595–3603
18. Shi, L., Zhang, Y., Cheng, J., Lu, H.: Skeleton-based action recognition with directed graph neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 7912–7921

19. Li, B., Zhu, C., Li, S., Zhu, T.: Identifying emotions from non-contact gaits information based on microsoft kinects. IEEE Transactions on Affective Computing **9** (2016) 585–591
20. Quiroz, J.C., Geangu, E., Yong, M.H.: Emotion recognition using smart watch sensor data: Mixed-design study. JMIR mental health **5** (2018) e10153
21. Ahmed, F., Sieu, B., Gavrilova, M.L.: Score and rank-level fusion for emotion recognition using genetic algorithm. In: 2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), IEEE (2018) 46–53
22. Zhang, Z., Song, Y., Cui, L., Liu, X., Zhu, T.: Emotion recognition based on customized smart bracelet with built-in accelerometer. PeerJ **4** (2016) e2258
23. Venture, G., Kadone, H., Zhang, T., Grèzes, J., Berthoz, A., Hicheur, H.: Recognizing emotions conveyed by human gait. International Journal of Social Robotics **6** (2014) 621–632
24. Karg, M., Kühnlenz, K., Buss, M.: Recognition of affect based on gait patterns. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **40** (2010) 1050–1061
25. Daoudi, M., Berretti, S., Pala, P., Delevoye, Y., Del Bimbo, A.: Emotion recognition by body movement representation on the manifold of symmetric positive definite matrices. In: International Conference on Image Analysis and Processing, Springer (2017) 550–560
26. Wang, W., Enescu, V., Sahli, H.: Adaptive real-time emotion recognition from body movements. ACM Transactions on Interactive Intelligent Systems (TiiS) **5** (2015) 1–21
27. Crenn, A., Khan, R.A., Meyer, A., Bouakaz, S.: Body expression recognition from animated 3d skeleton. In: 2016 International Conference on 3D Imaging (IC3D), IEEE (2016) 1–7
28. Randhavane, T., Bhattacharya, U., Kapsaskis, K., Gray, K., Bera, A., Manocha, D.: Identifying emotions from walking using affective and deep features. arXiv preprint arXiv:1906.11884 (2019)
29. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. arXiv preprint arXiv:1801.07455 (2018)
30. Shi, L., Zhang, Y., Cheng, J., Lu, H.: Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 12026–12035
31. Liu, Z., Zhang, H., Chen, Z., Wang, Z., Ouyang, W.: Disentangling and unifying graph convolutions for skeleton-based action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 143–152
32. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875 (2017)
33. Diao, Z., Wang, X., Zhang, D., Liu, Y., Xie, K., He, S.: Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence. Volume 33. (2019) 890–897
34. Yang, J., Zheng, W.S., Yang, Q., Chen, Y.C., Tian, Q.: Spatial-temporal graph convolutional network for video-based person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 3289–3299
35. Zhao, M., Zhong, S., Fu, X., Tang, B., Pecht, M.: Deep residual shrinkage networks for fault diagnosis. IEEE Transactions on Industrial Informatics **16** (2019) 4681–4690

36. Habibie, I., Holden, D., Schwarz, J., Yearsley, J., Komura, T.: A recurrent variational autoencoder for human motion synthesis. In: 28th British Machine Vision Conference. (2017)