2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Dance Dissected: Enhancing Labanotation Generation through Part-specific Attention with Transformers

Min Li, Jing Sang, Lina Du

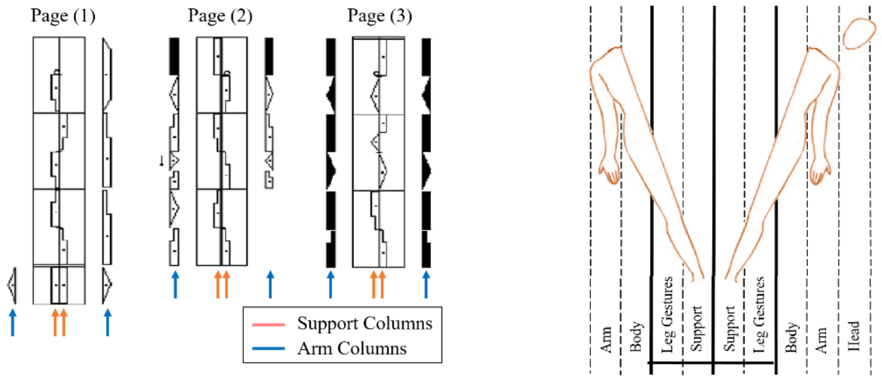
North China University of Technology, Beijing, China
limin@ncut.edu.cn

Abstract. Labanotation is a widely-used notation system for recording human dance movements. However, writing standard Labanotation scores requires extensive professional training. Automatically generating Labanotation scores from motion capture data can significantly reduce manual efforts in dance documentation and can serve as a powerful tool for preserving folk dances, in the protection work of intangible cultural heritages. Despite this, existing methods for Labanotation generation face challenges in capturing the more fluid and variable limb movements inherent in folk dance performances. In this paper, we introduce a novel Transformer-based model called PSA-Transformer (Transformer with Part-Specific Attention) to achieve more accurate Labanotation generation. First, we develop a Part-Specific Attention (PSA) module that adheres to the body part division rules of Labanotation. This module extracts spatial attention features at individual body part levels, enhancing the precision of movement capture. Then, this attention mechanism is integrated into an encoder-decoder architecture, enabling the model to learn global temporal dependencies within the feature sequences produced by the PSA module. As such, we sequentially generate corresponding Laban symbols using the decoder component of the PSA-Transformer. Extensive experiments on two real-world datasets demonstrate that our proposed model performs favorable against current state-of-the-art methods in automatic Labanotation generation.

Keywords: Labanotation generation, Part-Specific Attention, Transformer, limb movements, encoder-decoder.

1 Introduction

Labanotation is a common system for recording dance, used in teaching, sharing, and preserving dances [1, 2]. It's a good way to protect traditional folk dances, which are part of our cultural heritage. While methods like written descriptions, pictures, photos, and videos are used to document dances, Labanotation is accurate, efficient, easy to store, and useful for both teaching and communication. Several computer tools have been created to help write Labanotation scores [3–5]. However, creating these scores by hand is still a slow process that requires skilled professionals. Developing ways to automatically generate Labanotation scores could significantly help preserve traditional folk dances.



(a) Score example

(b) Columns and body parts

Fig. 1. Illustration of the Labanotation system. (a) Labanotation score example. (b) The corresponding relationship between columns of the Labanotation score and human body parts.

One common approach to automatic Labanotation generation is to identify Laban symbols from motion capture data and then arrange them into a score in the correct order. Motion capture devices record movement and produce a series of skeletal data frames. Unlike general action recognition which labels a series of movements with a single action [6–8], generating Labanotation involves creating a sequence of symbol labels from dance movements, where each movement of every body part is represented by a specific Laban symbol [1].

To understand our work, let’s briefly explain the Labanotation system. Figure 1(a) shows an example of a Labanotation score, which uses a three-line staff where symbols are written in columns. Symbols are arranged from bottom to top on each page, following the order of the dance. Each column represents a body part, and each symbol describes a movement of that part. For example, Figure 1(a) shows the movements of the legs in the support columns and the arms in the arm columns. Figure 1(b) illustrates how the columns relate to different body parts. The support and arm columns are the most used. Specifically, the support columns show how the body’s center of gravity moves during a step, which is essential in dance. Since it needs to accurately capture every detail of a dancer’s movements, Labanotation uses many different symbols.

Therefore, generating a correct Labanotation score requires accurate recognition of every atomic movement of each body part and proper assignment of signs according to the Labanotation writing conventions. Among the multiple Labanotation symbols, those indicating direction in the support columns are particularly important, as they describe the footwork, which is a fundamental part of dance. There are 24 direction symbols used in the support columns to represent movement on both sides of the body.

Many studies have focused on automatically generating Labanotation for the support columns. Early methods for converting motion capture data into Labanotation relied on handcrafted rules, but were limited in application. Subsequent research explored two-stage approaches with manual or automated movement segmentation followed by recognition using statistical learning or deep neural networks (including HMMs, Bi-

LSTMs, and GRU recurrent networks). More recently, one-stage methods employing Connectionist Temporal Classification (CTC) or sequence-to-sequence learning with feature extraction modules, including CNNs, CRNNs, or graph networks, have streamlined the process and avoided segmentation errors. However, accurately capturing the nuances of complex and variable dance movements (particularly in folk dances), remains a challenge. Existing feature extraction networks often focus on local joint relationships and limited temporal ranges, potentially missing crucial information for representing broader limb movements needed for precise Labanotation.

In this paper, we propose a novel PSA-Transformer model (Transformer with Part-Specific Attention networks). To describe flexible limb movements in detail and generate accurate Laban symbols, first, we propose a PSA module and employ the multi-head attention (MHA) mechanism to learn the feature correlations between joints on each body part defined in the Labanotation system. The body-part features are further aggregated to obtain an overall spatial feature for each sampled frame. Then, we apply the MHA mechanism in an encoder-decoder architecture to learn long-term global temporal dependencies in the output feature sequences of PSA module. Finally, the PSA-Transformer's decoder sequentially outputs a sequence of Laban symbols, effectively translating motion capture data into Labanotation scores for dances. Evaluations on two large datasets demonstrate that our model achieves state-of-the-art results, surpassing the performance of existing one-stage and two-stage methods.

2 Related Work

Numerous methods have been developed to tackle the problem of automatic Labanotation generation. Researchers focus on the Laban symbols in support columns, and the proposed methods are usually based on motion capturing technique. First, precise human motion data of dances are recorded via professional motion capturing devices. Then, spatial analysis, either using specially designed rules or machine learning methods, are employed to relate every single movement of each body part to a Laban directional symbol. Finally, a complete Labanotation score is constructed under the guidance of domain knowledge of music and dance.

Early approaches based on Spatio-temporal movement analysis [9–13] generally rely on artificially designed rules to translate motion capture data into Laban symbols frame-by-frame. However, the applicability of these methods is limited by the constraints of their handcrafted rules.

To overcome this limitation, subsequent researches introduce two-stage methods that first manually segment movements, then recognize these segments using conventional statistical learning techniques or deep neural networks. Recognition algorithms employed in this stage include hidden Markov models [14, 15], Extreme Learning Machine [16], bidirectional Long Short Term Memory (BiLSTM) [17], bidirectional Gated Recurrent Unit (BiGRU) [18], and combinations of LSTM and LieNet networks [19]. More recently, to reduce manual effort, some frameworks [20, 21] have incorporated auto-segmentation algorithms. However, errors in the segmentation phase can significantly degrade the quality of Labanotation generation.

To further refine Labanotation generation and eliminate segmentation errors, one-stage methods [22–28] have been proposed. Xie et al. [23, 24] utilize Connectionist Temporal Classification (CTC) [29] to directly convert per-frame skeletal motion features into sequences of Laban symbols. The features are extracted using Convolutional Recurrent Neural Networks (CRNNs) or Graph Neural Networks (GNNs). Li et al. developed a series of sequence-to-sequence (seq2seq) learning methods based on recurrent neural networks (RNNs) to generate Labanotation scores from continuous motion capture data, utilizing skeletal motion features extracted via CNNs [22], CRNNs [28], Gesture-Sensitive graph convolutional networks (GS-GCNs) [26],or Multi-scale graph attention networks (MS-GATs) [27].

While these one-stage methods offer a streamlined solution and avoid segmentation errors, Labanotation generation performance remains a challenge, particularly with the complexity and fluidity of dance movements. Particularly, folk dances often exhibit diverse limb movements and variations between performers, making precise capture and representation difficult. The Labanotation system demands fine-grained analysis of joint movements, and existing feature extraction networks like CRNN [22, 23], GS-GCN [26] and MS-GAT [27] often focus on local joint relationships and limited temporal ranges, potentially overlooking crucial information for capturing broader limb-level movements.

In this paper, we deal with this problem by proposing a transformer-based approach with a novel Part-Specific Attention network module. Detailed method description is as follows.

3 Methodology

3.1 Overview

This work aims to generate sequences of corresponding Laban direction symbols from motion capture data, specifically focusing on the support columns of movement. The overall process of our proposed PSA-Transformer is illustrated in Figure 2. Initially, we employ a motion data preprocessing strategy consistent with the approach described in [26], including coordinate transformation and normalization. For each frame, the proposed PSA module leverages the attention mechanism to extract spatial features based on body parts, allowing us to better model the correlations between joints and capture the nuances of flexible limb movements in dance. These spatial features are then aggregated to create a comprehensive representation of the frame. The resulting feature sequence, $\{g_1, g_2, \ldots, g_L\}$, is then positionally encoded and input into the temporal encoder-decoder of the Transformer, producing a sequence $\{z_1, z_2, \ldots, z_L\}$. This encoder-decoder architecture enables the learning of dynamic characteristics of body movements throughout the dance performance. Finally, the decoder outputs a sequential series of Laban symbols. By arranging these generated symbols along the dance timeline, a complete Labanotation score for the dance can be created, effectively achieving automatic Labanotation score generation from motion capture data. A detailed explanation of our proposed method is as follows.

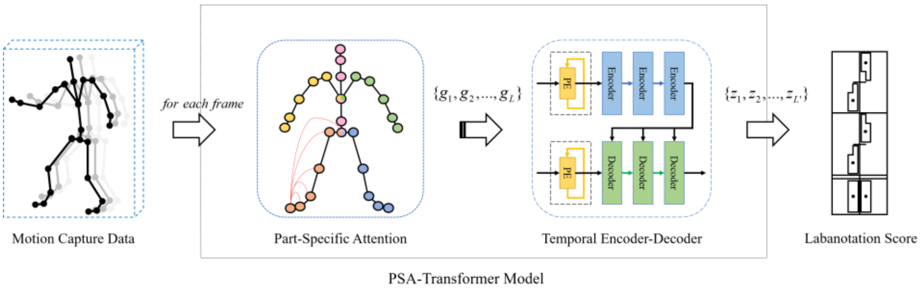


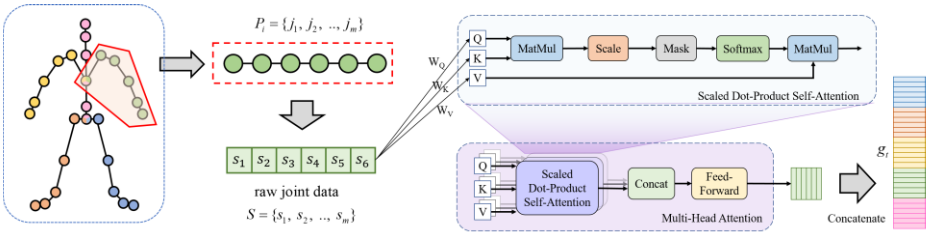
Fig. 2. Illustration of the overall method. The preprocessed motion capture data sequence is fed into the PSA-Transformer. For each frame, the proposed PSA module extracts spatial features by body parts via attention mechanism. In the figure, the color of joints differentiates body parts, and the red lines starting from Joint *right-foot* depicts the learned attention weights relative to other joints in a body part, as an example. Then, the outputs of PSA module form a sequence $\{g_1,\ g_2, \ldots,\ g_L\}$. We apply positional encoding and feed the spatial feature sequence to the encoder-decoder architecture to learn temporal dependencies across different frames over time and get a sequence $\{z_1,\ z_2, \ldots,\ z_L\}$. Finally, the Laban symbols are sequentially generated by the decoder, thus a Labanotation score is composed.

3.2 Part-specific Attention Module

Our method begins with a novel Part-Specific Attention (PSA) module specially designed to capture the characteristics of flexible limb movements. This PSA module learns correlations between joints within each body part, and enhances these joint features using the multi-head attention (MHA) mechanism. These enhanced joint features are then aggregated to form the overall spatial feature representation of the body skeleton for each frame.

We begin by explaining the criteria for dividing the body into parts, and then introduce the subsequent process of extracting features specific to each part. Guided by the principles of Labanotation, which separately records the movements of each limb and the trunk with directional symbols, we divide the motion capture data into distinct body parts based on the skeletal hierarchy. As illustrated in Figure 3, each part is visually distinguished using different colors. Each body part contains several joint points. Denoting a body part as $\mathbf{P}$, we can break down a frame of motion capture data as a group of body parts $\{P_1, P_2, .., P_n\}$, where we set $n = 5$ considering the structure of body skeletal data. Each body part consists of several joints and can be seen as a joint sequence $\{j_1, j_2, .. j_m\}$ based on spatial hierarchical relationships, as shown in Figure 3. For simplicity, we maintain a consistent number of joints, m, across all body parts.

For each body part $\mathbf{P}$, we extract spatial feature from the raw joint data using the multi-head attention (MHA) mechanism, MHA explicitly models relationships between all states within the sequence, enabling the learning of global dependencies between sequential joint positions.



$P_i = \{j_1, j_2, ..., j_m\}$
s_1 s_2 s_3 s_4 s_5 s_6
raw joint data
$S = \{s_1, s_2, ..., s_m\}$
W_Q
W_K
W_V
Q
K
V
MatMul
Scale
Mask
Softmax
MatMul
Scaled Dot-Product Self-Attention
Scaled Dot-Product Self-Attention
Concat
Feed-Forward
Multi-Head Attention
g_t
Concatenate

Fig. 3. Illustration of the PSA module. Each part of human body, represented as $\mathbf{P}$, is processed as a sequence of joints based on skeletal hierarchical structure. For each part, the multi-head attention (MHA) mechanism is applied to extract spatial features from sequential raw joint data. The obtained features are concatenated to form a final matrix g_t at frame t, which constitutes the spatial feature representation for the current frame of motion capture data.

Specifically, we apply the scaled dot-product attention [30], as illustrated in the right part of Figure 3. There are three inputs for self-attention computation: queries, keys, and values, represented with $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$. To learn dependencies across the input sequence itself, the *self*-attention is applied. In this case, all the query, key, and value are computed based on the same input sequence. Denoting the matrix representation of the input sequence as $\mathbf{S} = \{s_1, s_2, \ldots, s_m\}$ where m is the length of sequence (which is also the consistent number of joints in each part), the three inputs $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ to the self-attention module can be obtained via three weight matrices:

$$\begin{aligned} \mathbf{Q} &= \mathbf{S}\mathbf{W}_Q, \\ \mathbf{K} &= \mathbf{S}\mathbf{W}_K, \\ \mathbf{V} &= \mathbf{S}\mathbf{W}_V, \end{aligned} \tag{1}$$

where $\mathbf{S}$ is of dimension $m \times d_{model}$, $\mathbf{W}_Q$ is of dimension $d_{model} \times d_Q$, $\mathbf{W}_K$ of $d_{model} \times d_K$, and $\mathbf{W}_V$ of $d_{model} \times d_V$. We set $d_Q = d_K = d_V = d_{model} / n_{head}$, where n_{head} is the number of parallel heads in multi-head attention. Thus, the attention weight a^{self} is computed by the dot-product of queries $\mathbf{Q}$ and keys $\mathbf{K}$, following with a normalization with a scalar $\sqrt{d_V}$ and the softmax function, as follows:

$$a^{self} = \text{softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_V}})\mathbf{V}. \tag{2}$$

Here, the subscript *self* indicates that a^{self} is a self-attention weight. T denotes the transpose operation. It is easy to know that the dimension of a^{self} is $m \times d_V$.

The computation above allows each state within the input sequence to be represented with weighted contributions. Additionally, during decoding along the time axis, self-attention requires a mask to prevent the model from accessing future information when computing attention weights at each step t.

Then, to improve the model's ability to represent complex relationships, the attention computation is performed in parallel n_{head} times, each with different learned parameters. The outputs of these parallel attention heads are then concatenated to form the multi-head attention output:

$$a^{mhself} = \mathrm{concat}(a_1^{self},\ a_2^{self},\ldots,\ a_{n_{head}}^{self})\mathbf{W}_O, \tag{3}$$

where $\mathbf{W}_O$ is a weight matrix of dimension $d_V n_{head} \times d_{model}$, representing a fully-connected feed-forward network. a^{mhself} denotes the output of multi-head self-attention module, of dimension $m \times d_{model}$. The MHA computation procedure is depicted in the lower middle part of Figure 3.

More specifically, for the feature extraction on a body part $P_i(i \in \{1, 2, .., n\})$, the length of input sequence is the joint number m, and feature of P_i (i.e. the output of MHA) is in shape $m \times d_{model}$. Finally, we further concatenate the features of all body parts to form a matrix in $n \times m \times d_{model}$ as the spatial feature of current frame. By arranging all the frames in temporal order, we can obtain a sequence of spatial features, denoted as $\{g_1,\ g_2,\ldots,\ g_t,\ldots,\ g_L\}$, where g_t represents the frame feature, and L is the length of frame sequence.

3.3 Encoder-Decoder Network

In the proposed method, the PSA module takes in motion capture data sequences and outputs corresponding sequences of motion features. To learn the temporal dependencies in the motion feature sequences from PSA, we apply an encoder-decoder network to learn and transform the motion feature sequences to Laban symbol sequences.

The output spatial feature sequence of the PSA module, $\{g_1,\ g_2,\ldots, g_t,\ldots, g_L\}$, is the input of the encode-decoder network. We apply positional encoding to the spatial feature sequence, and denote the positionally encoded sequence as $\{x_1,\ x_2,\ldots, x_L\}$. Generally, the learning of temporal dynamic characteristics and global dependencies across frames over a long sequence is accomplished via the encode-decoder architecture. First, the encoder processes the input sequence to learn a compact representation, also utilizing the above-stated multi-head attention (MHA) mechanism for this purpose. Denoting with f_{encode}, we can describe the encoding procedure as:

$$\mathbf{E} = f_{encode}(\{x_1,\ x_2,\ldots,\ x_t\}), \tag{4}$$

where $\{x_1,\ x_2,\ldots,\ x_t\}$ is the positional-encoded sequence, and $\mathbf{E}$ is the encoded output. Then, we describe the decoding procedure as:

$$\mathbf{Z} = f_{decode}(\mathbf{E}, \{< BOS >\}), \tag{5}$$

where $\mathbf{Z}$ denotes the decoded sequence output, and $< BOS >$ represents *begin-of-sequence*. Using a greedy strategy that outputs the label with the biggest probability, the decoder generates a sequence of Laban symbols, which is the output sequence denoted as $\mathbf{Z} = \{z_1, z_2, \ldots, z_t, \ldots, z_{L'}\}$, where z_t denotes the Laban symbol at time z_t, and L' is the length of the output sequence. Finally, the generated Laban symbols are organized into Labanotation scores along with the time order.

4 Experiments

We assess the proposed PSA-Transformer model using the datasets *LabanSeq16* and *LabanSeq48*, which are two common benchmarks for skeletal motion sequence analysis. Our method demonstrates competitive performance compared to both two-stage (automatic segmentation-based) and one-stage approaches. We also include parameter sensitivity analysis results, detailed below.

4.1 Experimental Setup

Datasets and Evaluation Metric. *LabanSeq16* and *LabanSeq48* are two commonly used datasets for automatic Labanotation generation, captured using the OptiTrack$^{\text{TM}}$ motion capture system [31] motion capture system. We divided each dataset using a portion of 50% for training and the remainder for testing. Performance was evaluated using Laban symbol accuracy, consistent with previous work [22, 25, 26].

Implementation Details. In the proposed model, for PSA, we use part number $n = 5$ and joint number of each part $m = 6$. In the MHA of the PSA module, we use 2 heads and 2 stacked encoders. The d_{model} of spatial MHA is 32. For the Encoder-Decoder network, we use 4 encoders followed by 4 decoders, and the head number is 4. The d_{model} of temporal MHA is 256. The whole PSA Transformer model is implemented with PyTorch. All the experiments are conducted on a workstation with an RTX 3090 GPU.

4.2 Experimental Results and discussions

Comparison with existing methods. We benchmarked our method against existing Labanotation generation techniques, including both two-stage and one-stage approaches. Table 1 presents a comparison of symbol generation accuracies across the two datasets. For the two-stage methods, we utilized the automatic segmentation technique from [21]. Results for existing methods on *LabanSeq16* were taken directly from published work; for *LabanSeq48*, we obtained results using the authors' provided code (as some prior studies did not include this dataset in their evaluations). The final row of Table 1 showcases the performance of our proposed PSA-Transformer.

Table 1. Comparison of different Labanotation generating methods on symbol generation accuracy (%). The PSA-transformer model obtains the best results.

Methods	LabanSeq16	LabanSeq48
Auto Seg + HMM [15]	74.17	69.13
Auto-Seg + LSTM [17]	76.57	71.28
Auto-Seg + GRU [18]	77.22	73.21
Auto-Seg + C-GRU [21]	78.92	73.96
CRNN-CTC [23]	72.80	68.92
Seq2seq [22]	73.60	70.89
DFGNN-CTC [24]	87.79	82.02
GS-GCN + RA-Attention [26]	89.84	89.40
Fusion Feature + CRNN-based - Attention-seq2seq [28]	86.21	91.61
Baseline Transformer	93.16	92.24
LabanFormer [27]	94.44	94.86
PSA-Transformer (Proposed)	**94.93**	**95.20**

From Table 1, we can see that the proposed model obtains the best results on both datasets *LabanSeq16* and *LabanSeq48* comparing with both existing two-stage and one-stage methods. In addition, the transformer-based approaches generally show better results compared to previous methods, while the proposed method performs the best. This indicates that our PSA-Transformer model is more suitable for generating Labanotation scores from motion capture data. Moreover, we can see that in comparison to other transformer-based models with similar temporal encoder-decoder architectures, the proposed PSA-Transformer performs the best, which shows that the novel PSA module provides more fine-grained spatial features by aggregating joint features of body parts that can contribute to higher accuracies.

Discussion on spatial feature extraction methods. The proposed PSA module is based on the part-level feature extraction and aggregation. To demonstrate the benefit of part division, we compare the proposed PSA module with a non-part-specific approach. Specifically, we perform the same MHA operations directly on the whole skeleton of each data frame, without dividing into parts, and take the output as the spatial feature. We name this module as spatial-attention (SA). Other components of the pipeline are kept the same. In essence, this is a replacement of spatial feature extraction modules, and we conduct experiments to compare between them.

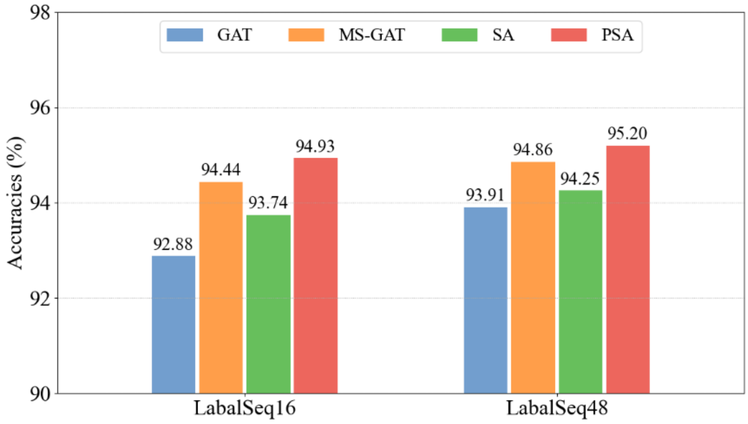


Fig. 4. Comparison of accuracies (%) of different spatial feature extraction modules. The GAT (graph attention network) and MS-GAT (multi-scale GAT) methods are from [27]. PSA is the proposed Part-Specific Attention network, while SA is a comparison method that uses the same MHA modules with PSA, but does not perform part-specific feature extraction.

In addition, we note that the LabanFormer model proposed in [27] also employed transformer to sequentially generate Laban symbols. A main difference of our method to LabanFormer is the spatial feature extraction module. LabanFormer [27] proposed a multi-scale graph attention network (MS-GAT) to learn spatial features in each frame of motion sequences, while in our work the PSA module does the similar thing. The GAT and MS-GAT modules from [27] are based on a kind of attention mechanism different from that in this work. Therefore, we conduct a horizontal comparison on different spatial feature extraction modules, including GAT, MS-GAT, the above-stated SA, and the proposed PSA. The experimental results are shown in Figure 4.

We evaluate the final Labanotation generation performance on both datasets of *LabanSeq16* and *LabanSeq48* for the four methods. All the compared networks share the same input data and the remaining encode-decoder structures use the same Transformers. From Figure 4, we can see that the proposed PSA module performs the best. The SA module, which simply extracts spatial feature using MHA from the whole skeleton, does not show advantages over MS-GAT. This indicates that the increase of accuracy mostly comes from the part-specific design. With the proposed PSA module, we can obtain more distinguishable human-body spatial features by learning at the level of body parts. The experimental results demonstrate the effectiveness of the proposed PSA module for the representation of flexible limb movements.

Qualitative results. Finally, we demonstrate an example of the final output Labanotation score with generated Laban symbols in the support column for a series of lower limb movements, as in Figure 5. The motion sequence contains four steps: forward (left foot), right-forward (right foot), right-forward (left foot), and forward (right foot). The generated Laban symbols written in the support columns accurately reproduce the above movements.

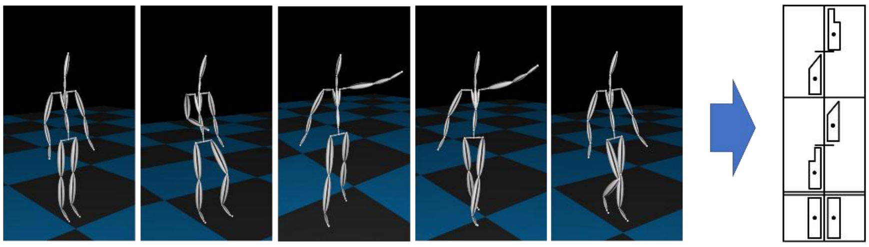


Fig. 5. An example of output Labanotation score composed of a few Laban symbols, which are generated from the motion capture data of several walking movements.

5 Conclusion

In this paper, we propose a new Transformer model based on novel Part-Specific Attention Networks (PSA-Transformer) to automatically generate Labanotation scores in support columns from motion capture data. The proposed PSA extracts and aggregates the features extracted at part-level to obtain a better representation of the flexible dance limb movements. With the self-attention mechanism of the Encoder-Decoder architecture, the temporal dependencies in the motion sequence are captured. Extensive experiments show that the proposed method can generate Labanotation scores at higher accuracy and outperform the state-of-the-art methods in the automatic Labanotation generation task.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. 61971007, and the Scientific Research Foundation of North China University of Technology.

References

1. Guest, A.H.: Labanotation: the system of analyzing and recording movement (4th ed.). Routledge, New York, NY, USA (2014).
2. Laban, R., Ullmann, L.: The mastery of movement, 4th ed. Dance Books Ltd., UK (2011).
3. Kojima, K., Hachimura, K., Nakamura, M.: LabanEditor: Graphical editor for dance notation. In: 11th IEEE International Workshop on Robot and Human Interactive Communication Proceedings. pp. 59–64 (2002). https://doi.org/10.1109/ROMAN.2002.1045598.
4. Hunt, F., Politis, G., Herbison-Evans, D.: Led & lintel: A Windows mini-editor and interprepter for Labanotation. Basser Department of Computer Science, University of Sidney, Sydney, NSW, Australia (2010).
5. Venable, L., Ralley, D., Alpert, V., Carol, F.: Laban Writer Workshop, (2004).

6. Lu, H., Wang, T.: Joint Spatiotemporal Collaborative Relationship Network for Skeleton-Based Action Recognition. In: 20th International Conference on Intelligent Computing (ICIC 2023). pp. 775–786. , Zhengzhou, China (2023).
7. Huang, J., Liu, X., Hu, H., Tang, S., Li, C., Zhao, S., Lin, Y., Wang, K., Liu, Z., Lian, S.: Spatial-Temporal Transformer Network for Continuous Action Recognition in Industrial Assembly. In: 20th International Conference on Intelligent Computing (ICIC 2024). pp. 114–130. , Tianjin, China (2024).
8. Qi, M., Liu, Z., Li, S., Zhao, W.: Skeleton Action Recognition Based on Spatial-Temporal Dynamic Topological Representation. In: 20th International Conference on Intelligent Computing (ICIC 2024). pp. 249–261. , Tianjin, China (2024).
9. Hachimura, K., Nakamura, M.: Method of generating coded description of human body motion from motion-captured data. In: IEEE International Workshop on Robot and Human Interactive Communication (2001).
10. Choensawat, W., Nakamura, M., Hachimura, K.: GenLaban: A tool for generating Labanotation from motion capture data. Multimed. Tools Appl. 74, 10823–10846 (2015).
11. Choensawat, W., Nakamura, M., Hachimura, K.: Applications for Recording and Generating Human Body Motion with Labanotation. Dance Not. Robot Motion. 391–416 (2016). https://doi.org/10.1007/978-3-319-25739-6_19.
12. Chen, H., Qian, G., James, J.: An autonomous dance scoring system using marker-based motion capture. In: IEEE Workshop on Multimedia Signal Processing (2005).
13. Guo, H., Miao, Z., Zhu, F., Zhang, G., Li, S.: Automatic labanotation generation based on human motion capture data. In: Chinese Conference on Pattern Recognition (2014).
14. Li, M., Miao, Z., Ma, C.: Automatic Labanotation Generation from Motion-Captured Data Based on Hidden Markov Models. In: IAPR Asian Conference on Pattern Recognition (ACPR) (2017).
15. Li, M., Miao, Z., Ma, C.: Dance Movement Learning for Labanotation Generation Based on Motion-Captured Data. IEEE Access. 7, 161561–161572 (2019).
16. Zhang, X., Miao, Z., Zhang, Q.: Automatic Generation of Labanotation Based On Extreme Learning Machine with Skeleton Topology Feature. In: IEEE International Conference on Signal Processing (ICSP) (2018).
17. Zhang, X., Miao, Z., Zhang, Q., Wang, J.: Skeleton-based automatic generation of Labanotation with neural networks. J. Electron. Imaging. 28, 1 (2019).
18. Hao, S., Miao, Z., Wang, J., Xu, W., Zhang, Q.: Labanotation Generation Based on Bidirectional Gated Recurrent Units with Joint and Line Features. In: IEEE International Conference on Image Processing (ICIP) (2019).
19. Xie, N., Miao, Z., Wang, J.: Skeleton-based Labanotation generation using multi-model aggregation. In: IAPR Asian Conference on Pattern Recognition (ACPR) (2019).
20. Wang, J., Miao, Z.: A method of automatically generating Labanotation from human motion capture data. In: International Conference on Pattern Recognition (ICPR) (2018).
21. Li, M., Miao, Z., Ma, C., Li, A., Zhou, T.: An Automatic Framework for Generating Labanotation Scores From Continuous Motion Capture Data. In: IEEE International Conference on Multimedia & Expo (ICME) (2020).
22. Li, M., Miao, Z., Ma, C.: Sequence-to-sequence Labanotation Generation based on Motion Capture Data. In: the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2020).

23. Xie, N., Miao, Z., Wang, J., Zhang, Q.: End-to-end Method for Labanotation Generation from Continuous Motion Capture Data. In: IEEE International Conference on Multimedia & Expo (ICME) (2020).
24. Xie, N., Miao, Z., Zhang, X.-P., Xu, W., Li, M., Wang, J.: Sequential Gesture Learning for Continuous Labanotation Generation Based on the Fusion of Graph Neural Networks. IEEE Trans. Circuits Syst. Video Technol. (2022).
25. Li, M., Miao, Z., Zhang, X.-P., Xu, W.: An Attention-seq2seq Model Based on CRNN Encoding for Automatic Labanotation Generation from Motion Capture Data. In: the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2021).
26. Li, M., Miao, Z., Zhang, X.-P., Xu, W., Ma, C., Xie, N.: Rhythm-Aware Sequence-to-Sequence Learning for Labanotation Generation With Gesture-Sensitive Graph Convolutional Encoding. IEEE Trans. Multimudias. (2021).
27. Li, M., Miao, Z., Lu, Y.: LabanFormer: Multi-scale graph attention network and transformer with gated recurrent positional encoding for labanotation generation. Neurocomputing. (2023).
28. Li, M., Miao, Z., Xu, W.: A CRNN-based attention-seq2seq model with fusion feature for automatic Labanotation generation1. Neurocomputing. 454, 430–440 (2021).
29. Graves, A., Jaitly, N.: Towards End-To-End Speech Recognition with Recurrent Neural Networks. In: International Conference on Machine Learning (ICML) (2014).
30. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need. ArXiv170603762 Cs. (2017).
31. OptiTrack: OptiTrack Motion Capture Systems for applications ranging from animation, virtual reality, movement sciences, robotics and more, https://www.optitrack.com/, (2021).