Research Article

A Motion Capture Data-Driven Automatic Labanotation Generation Model Using the Convolutional Neural Network Algorithm

Jiang Yao and Yang Chen

School of Music and Dance, Shenzhen University, Shenzhen 518060, China

Correspondence should be addressed to Yang Chen; yanischen@szu.edu.cn

Received 2 November 2021; Revised 23 November 2021; Accepted 4 February 2022; Published 23 February 2022

Academic Editor: Xin Ning

Copyright © 2022 Jiang Yao and Yang Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

All human movements can be effectively represented with labanotation, which is simple to read and preserve. However, manually recording the labanotation takes a long time, so figuring out how to use the labanotation to accurately and quickly record and preserve traditional dance movements is a key research question. An automatic labanotation generation algorithm based on DL (deep learning) is proposed in this study. The BVH file is first analyzed, and the data are then converted. On this foundation, a CNN (convolutional neural network) algorithm for generating the dance spectrum of human lower-limb movements is proposed, which is very good at learning action space information. The algorithm performs admirably in terms of classification and recognition. Finally, a spatial segmentation-based automatic labanotation generation algorithm is proposed. To begin, every frame of data is converted into a symbol sequence using spatial law, resulting in a very dense motion sequence. The motion sequence is then regulated according to the minimum beat of motion obtained through wavelet analysis. To arrive at the final result, the classifier is used to determine whether each symbol is reserved or not. As a result, we will be able to create more accurate dance music for simple human movements.

1. Introduction

Labanotation is a type of dance music created by Hungarian Laban that records human movements with various elements and symbols. Labanotation is made up of lines and symbols, just like music staff. It is widely used in recording and teaching dance movements [1] because it can accurately and conveniently record and analyze all human movements. It is indispensable for the dissemination and communication of culture [2]. As a result, using labanotation as a method of recording is extremely beneficial to the preservation and transmission of Chinese traditional folk dance.

In comparison to the turn of the century, western dance art theory has progressed significantly, with notable achievements in recording dance score, dance steps, and rhythm, the most influential of which is labanotation, which not only has a significant impact on dance art but is also

widely used in other fields involving human body movements [3, 4]. When motion capture technology first became available, it was quickly applied to the recording of dance movements. Computers have matured research on data compression and preservation, and motion capture technology converts three-dimensional human motion into binary data that can be recognized by computers. Using mathematical models to store human motion data, on the other hand, is not only expensive but also easily damaged, difficult to read, and publicize. Keeping records of national dance movements is thus an important research topic.

The main work of this study is to study an efficient algorithm for automatic generation of labanotation based on motion capture data. Firstly, according to the characteristics of human motion, the skeletal features that conform to the topological structure of human body are designed. An efficient DL (deep learning) algorithm [5] is proposed for the

first time to recognize human upper- and lower-limb movements. Finally, the designed features and upper- and lower-limb movements recognition algorithm are integrated into the automatic generation platform of dance spectrum, and an end-to-end automatic generation system from motion capture data to labanotation is realized.

The study's novelty is that the labanotation analysis of lower-limb movements focuses on the dynamic process of movements, whereas CNN (convolutional neural network) can capture the context in time-series data very well, thanks to its special gate control. As a result, this study examines and proposes the use of CNN to recognize lower-limb movements. The proposed CNN-based method's superiority in processing dynamic time-series data, as well as the high efficiency of lower- and middle-limb motion generation in labanotation, is demonstrated by experimental comparison and analysis of the proposed method and other recognition algorithms.

2. Related Work

The advancement of computer science has had an impact on the advancement of many fields [6–8]. Labanotation has also investigated the use of computers for drawing and displaying. Literature [9] used motion capture equipment to collect data from a 41-point model, emphatically analyzing movements such as rotation, pace, jumping, lifting, and descending, and the Laban Writer output dance music through the interface, so the movements studied lacked universality. Literature [10] analyzes the supporting and unsupported actions separately and proposes a method of automatically generating labanotation based on knowledge rules, which details each part of the method and realizes the platform of automatically generating labanotation, according to the labanotation analysis principle. The transformation process of labanotation was realized in literature [11], in which segmentation and recognition were completed automatically, but the segmentation and action recognition rates were low and unable to meet the actual application level. Literature [12] has perfected the conversion of dance spectrums. The new method is a rule-based dance spectrum conversion system, but it is only good for some basic action research. In light of inaccurate segmentation, literature [13] improves the segmentation algorithm and adopts an action segmentation algorithm based on rhythm analysis, which is only applicable to behaviors with a strong sense of rhythm, is complex, and cannot be applied in practice.

To automatically generate labanotation by computer, it is essential to have a technology that can collect human actions. Before the technology of motion capture matured, no one had ever formally studied the automatic generation of labanotation. Literature [14] puts forward the method of generating labanotation of upper-limb movements by mapping labanotation symbols in three-dimensional coordinate space, which is a spatial analysis based on Laban notation principle. However, in their research, the expression of labanotation in upper-limb movements was emphatically analyzed, while the more complex lower-limb movements and the movement of human center of gravity

were not analyzed and discussed. Literature [15] analyzes the motion based on motion capture data by a rule-based method, but when generating labanotation, the pause and segmentation of the motion are not considered, so the generated labanotation is of low quality. Literature [16] summarizes the above experience and perfects the whole automatic dance spectrum conversion process, but its recognition rate is very low and far from the application level. Literature [17] optimized it and proposed a rule-based dance spectrum transformation system. However, there is still a long way to go before the practical teaching level.

3. Overview of Labanotation and Motion Capture

3.1. Introduction to Labanotation. Labanotation is composed of a spectrum table and symbols of movement elements. First of all, in music, we can see that every syllable is expressed in the form of musical symbols, and the style of dance movements has a great relationship with beats. Laban proposed to use rectangular symbols to represent each movement beat, and the length of symbols represents the duration of movements. Secondly, in music, it can be seen that a piece of music is recorded in a horizontal way, while Laban analyzes the movement recording process of dance movements in time and limbs, and proposes to record a piece of dance movements by means of column longitudinal recording.

Labanotation is based on science and vividly restores all human activities [18]. His influence on the communication and dissemination of different dances in different countries is obvious. At present, labanotation has become a common language in dance circles.

In eleven vertical columns, labanotation records the movements of the main parts of the human body. These 11 columns are arranged on both sides of the center line and are used to describe the movements of various body parts. The left half of the human body and the head are recorded and described using columns on the left side of the center line; the right half of the human body and the head are recorded and described using columns on the right side of the center line. The basic principles of human anatomy, as well as the characteristics of standing and walking, are well represented by this notation method. More importantly, it can make it easier for people to read dance music and comprehend the descriptions of each movement in each section.

3.2. The Principle of Labanotation Drawing Based on Motion Capture Data. The automatic drawing of labanotation is realized by processing motion capture data. Each dance symbol in labanotation corresponds to an action, while each frame of data in BVH file contains the movement of 26 joint points of human body. Therefore, in the process of automatically generating labanotation based on motion capture data, BVH files need to be analyzed and processed, as shown in Figure 1.

Firstly, the BVH file is parsed, the joint semantics are uniquely determined by quadruple, and the format of the data area file is converted into position data.

FIGURE 1: Generating labanotation schematic diagram from motion capture data.

Then, the motion capture data representing human motion are segmented and extracted, and four kinds of segmentation algorithms are adopted based on speed threshold and its improved algorithm, PCA (principal component analysis) and its extended algorithm, beat algorithm, and feature coupling algorithm, so that the whole human motion is segmented into basic element actions [19, 20].

Then, the segmented data are processed by attitude analysis. The posture transformation of the supported and unsupported parts is analyzed. Finally, the human motion transformation is recorded in the form of dance notation.

4. Research Method

On the basis of the dynamic characteristics of lower-limb movements in dance music records, this study proposes for the first time to use CNN model in DL algorithm to generate lower-limb movements dance music based on motion capture data. At the same time, referring to Laban notation rules, this study puts forward a spatial segmentation method for upper-limb movement recognition.

4.1. Preprocessing of 3D Motion Capture Data. Considering the research difficulties in this study, it is particularly important to extract effective motion features from experimental video data, and the quality of feature selection will directly affect the computation, recognition accuracy, and algorithm robustness of the platform, as shown in Figure 2.

As shown above, the appropriate motion features are selected, but they cannot be directly calculated from the original BVH data, so it is necessary to convert the feature space, that is, convert the BVH format file [21].

The concrete idea is to hypothesis that there is a point (x_0, y_0, z_0) and transform coordinate system to obtain an inertial coordinate system (x_1, y_1, z_1) , as shown in the following equation:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = R \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}, \quad (1)$$

where R is the rotation matrix to be constructed between two points, and the data in the experiment are rotated around ZXY axis in turn, then R and the matrix obtained along XYZ axis can be obtained as shown in the following equation:

$$R = R_z R_x R_y, \quad (2)$$

where R_z represents the rotation matrix obtained along the Z axis, and R_x, R_y can be obtained in the same way. Rotation matrix is shown as follows:

$$\begin{aligned} R_z(\theta) &= \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ R_x(\theta) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}, \\ R_y(\theta) &= \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}. \end{aligned} \quad (3)$$

According to the above formula, the transformation from object coordinate system (x_0, y_0, z_0) to inertial coordinate system (x_1, y_1, z_1) is completed, and then the world coordinate system (x_2, y_2, z_2) is obtained by translation.

Because every adjacent node of BVH file is related, the position of the child node can be obtained by the position of the parent node. Assuming that the inertial rotation matrix of all previous nodes of node J is $R_1, R_2, \dots, R_r$, the offset of this node relative to its predecessor nodes in inertial coordinate system is shown in the following equation:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = R_r \left(R_r \left(\dots \left(R_2 \left(\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right) \right) \right) \right). \quad (4)$$

Assuming that the world coordinate system of the node at the previous moment is (x_0, y_0, z_0) and the world coordinate system of the node is (x_1, y_1, z_1) , after ten frames, the displacement, velocity, and acceleration are shown as follows:

$$s = \sqrt{(x_1 + x_0)^2 + (y_1 + y_0)^2 + (z_1 + z_0)^2}, \quad (5)$$

$$v = \frac{s}{Frames \times framesTime}, \quad (6)$$

$$a = \frac{v_2 - v_1}{Frames \times framesTime}. \quad (7)$$

Here, $Frames, framesTime$ exist in the data block of BVH file and can be read and used directly. Displacement,

FIGURE 2: Body pose feature parameters.

velocity, and acceleration can reflect the posture information of human body, and angle can reflect the rotation information of human body. The included angle of knee joint is taken as an example (the included angle of elbow joint is the same), as shown in Figure 3:

If the *leftUpLeg*-node world coordinate system is (x_0, y_0, z_0) , *leftLowLeg*-node world coordinate system is (x_1, y_1, z_1) , and *leftFoot*-node world coordinate system is (x_2, y_2, z_2) , the side length of the triangle formed by these three nodes of the human leg and the included angle of the knee joint can be calculated as follows:

$$a = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2}, \quad (8)$$

$$b = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}, \quad (9)$$

$$c = \sqrt{(x_0 - x_2)^2 + (y_0 - y_2)^2 + (z_0 - z_2)^2}, \quad (10)$$

$$\theta = \arccos\left(\frac{(a^2 + b^2 - c^2)}{2ab}\right). \quad (11)$$

4.2. Automatic Generation Algorithm of Labanotation Based on DL

4.2.1. Generation of Lower-Limb Labanotation. For the dynamic changes of lower-limb movements in dance music records, this study proposes for the first time to use the very powerful DL algorithm-CNN (convolutional neural network). The biggest feature of CNN is its sharing mechanism and unique convolution operation. The network can fully learn the spatial information, thus achieving the classification and recognition effect.

The main feature of CNN is that its network structure has different convolution calculations from other networks. It is one of the current representative algorithms of depth science. It is used in image processing [22], target detection [23], behavior recognition [24], and natural language processing and other fields [25–27] and has made extraordinary achievements [28].

On the whole, CNN is similar to other neural network structures that are composed of input layer, hidden layer, and output layer.

FIGURE 3: Angle of knee joint.

CNN's unique convolution kernel can learn the spatial information of human body movements well, and different convolution kernels can capture more information. Therefore, compared with traditional machine-learning methods, CNN's structure is more comprehensive in learning and representing tasks and can achieve better performance than traditional methods.

In this study, a CNN-based method for human motion recognition in labanotation is proposed for the first time. Figure 4 is the framework diagram of human motion recognition based on CNN proposed in this study.

The dimension of each input sample after extracting the vector features of human skeleton nodes from motion capture data is $20 * 14$, which is smaller than the input dimension of complex global human behavior recognition, so the CNN network structure proposed in this study has also changed. The entire network is optimized in terms of recognition accuracy and response time. The CNN structure proposed in this study is not particularly deep in hierarchy, as shown in Figure 4, and it is well suited to the task of automatic labanotation generation in recognition performance and time performance.

The CNN-based method for generating dance spectrum of lower limbs is proposed in this study, in which the network structure consists of two convolution layers, a pool layer, three fully connected layers, the output of the softmax function is the probability value of the human action category, and finally, the category with the largest probability value is taken as the result of network prediction [29].

FIGURE 4: Generation structure of lower-limb movement dance spectrum based on CNN.

In order to prevent overfitting, a dropout layer is connected behind each fully connected layer. The whole network uses random gradient descent algorithm to optimize the network and accelerate convergence. The objective function of the whole network is to minimize the cross loss entropy function [30]:

$$L_k - \sum_j t_{k,j} \log(p_{k,j}). \quad (12)$$

Here, $t_{k,j}$ indicates that the true label category of sample k is j , and $p_{k,j}$ indicates the predicted category of sample k by the network.

4.2.2. Upper-Limb Motion Segmentation and Labanotation Generation. When he raises his hand forward, he records a forward symbol, but when the hand returns to its original position, he does not record the backward symbol like the foot but records the original symbol. It can be seen that the recognition of arm movements is related to the current spatial position of the arm.

When Laban notation records upper-limb movements, it is judged according to the spatial position of the current motion state, and symbol changes occur when spatial changes are made. Based on this rule, this study proposes a spatial segmentation method.

The first step is to determine the symbol level, that is, the type of vertical direction, after obtaining the motion vector of each frame of arm movement. Because of the difference in people's height, a standing arm movement and a squatting arm movement may also be the same symbol, so the judgment in the vertical direction still needs to be obtained by using the vector included angle.

The motion vector is represented by $\vec{d}_1$, and the three-dimensional coordinates of the object coordinate system are represented by $\vec{x}, \vec{y}, \vec{z}$. The three vectors $\vec{x}, \vec{y}, \vec{z}$ are perpendicular to each other, so each axis can be regarded as

the normal vector of the plane determined by the other two axes, the projection vectors on the plane can be calculated from the vector $\vec{d}_1$ and the plane normal vector, and the projection vectors of the counter vector $\vec{d}_1$ on the three planes are $\vec{d}_x, \vec{d}_y, \vec{d}_z$, respectively. There are

$$\theta_{x,y,z} = \arccos \left(\frac{|\vec{x}, \vec{y}, \vec{z} \cdot \vec{d}_{x,y,z}|}{|\vec{x}, \vec{y}, \vec{z}| |\vec{d}_{x,y,z}|} \right). \quad (13)$$

The angle between vector d and three axes of object coordinate system can be calculated respectively. The forward, left or right front, left or right, left or right back, and backward motions can be judged by the included angle between the motion vector and the z axis, and whether it is left or right can be further determined by judging the included angle between the x axes, so that a horizontal symbol can be determined.

The dense motion sequence obtained in the second step. The result has the characteristics of Laban symbol recording dance, but every frame of data is converted, and the sequence obtained is particularly dense. The next step is to integrate these dense symbol sequences.

The third step is to integrate sequences based on the beat method. Both music and dance movements are inseparable from the beat. As long as you get the minimum beat of a dance, you can use this minimum beat to regulate the signal. Because the dance is not chaotic, you can think that they are all carried out on the basis of this minimum beat, that is, the beat of the whole dance movement will be an integral multiple of this minimum beat.

In this study, SVM (support vector machine) is used for classification judgment. SVM classifier is to find a hyper-plane to separate positive and negative samples, and its process can be understood as maximizing the minimum interval. The objective function is shown in the following formula:

$$\arg \max \left\{ \min(\text{lab}(w^T X + b)) \cdot \frac{1}{\|w\|} \right\}. \quad (14)$$

Here, lab represents the sample category, and $w^T X + b$ represents the hyperplane equation. The final result can be obtained through continuous optimization and iteration.

Figure 5 is a speed frame diagram of an arm rotating at a constant speed. It can be seen that different motions can be well distinguished by using angular velocity components in different directions. After obtaining the classifier, the Laban symbol sequence obtained in the previous step to the classifier is sent, and whether each symbol is retained or deleted is judged, so as to obtain the final accurate result.

5. Result Analysis and Discussion

5.1. Algorithm Analysis of Lower-Limb Movement Labanotation Generation. In this study, the proposed CNN-based lower-limb motion recognition method is verified on two published data sets of human lower-limb motion capture. Firstly, these two data sets are introduced as follows:

- (1) Data set I: The data set contains 16 types of lower-limb movements, and there are 400 action samples in each category, that is to say, the total number of samples in the data set is 6400. The length of each sample is 20 frames, and finally, the whole data set has 128,000 frames.
- (2) Data set II: The data set is established and expanded on the basis of data set I in this study, which enriches the low-order and high-order actions in labanotation, and contains 48 kinds of human actions that are consistent with the sample distribution in data set A and all adopt uniform distribution, thus eliminating the problem of disbelief in recognition accuracy that may be caused by the unbalanced distribution of sample categories. The final data set contains 19,500 samples and 379,000 frames.

The size of the convolution has a clear impact on image processing. The enhancement effect of action information interested in the image is not obvious if the convolution kernel used is small, but it may increase noise in the same frequency band and obscure some of the image's original details. If the convolution kernel used is large, the computation of convolution will increase as well, resulting in a high level of computational complexity. As a result, choosing the right convolution kernel size for the entire video image is crucial.

According to a large number of 2D CNN applications in human motion recognition, the convolution network structure with the convolution kernel size of 33 has the best effect, as shown in Figure 6. The convolution kernel of a 3D CNN has a time dimension of three in general, so the network structure proposed in this study uses a 3D convolution kernel with a size of 33.

The CNN structure proposed in this study can be effective for both data set I and data set II, which shows that the network structure has high robustness. Figures 7 and 8

FIGURE 5: Components of velocity in all directions.

FIGURE 6: Relationship between segmentation accuracy and sliding window size of convolution kernel.

show the accuracy of different algorithms in human motion recognition.

Compared with the traditional template matching method in reference [22], the recognition accuracy of the proposed CNN-based method on data set I is about 13% higher and about 5% higher than that of the method in reference [23].

Even on the data set II containing more action types and more action samples, the proposed method still achieves the highest recognition accuracy, with an average recognition rate of 95.58%, which is better than the previous research methods, and fully demonstrates the advantages of CNN in

FIGURE 7: Comparison of experimental results of upper- and lower-limb movements in data set I.

FIGURE 8: Comparison of experimental results of upper and lower-limb movements in data set II.

feature learning and representation compared with traditional machine-learning methods.

Compared with the recognition method in reference [24], the CNN-based method proposed in this study learns more spatial information, so it achieves better results. Even when the data samples are enlarged, CNN has achieved very good performance under the condition that the network structure and parameter settings are completely unchanged, which proves that the method based on CNN is efficient and robust in learning human motion space information.

Figures 9 and 10 show the time performance comparison between the proposed lower-limb motion recognition method based on CNN and other methods.

It can be seen that the method based on CNN is very fast, and it only takes 73 seconds on the data set I containing 6400 samples, while the method in reference [22] takes 382 seconds. However, the method in reference [23] takes longer, about 2000 seconds, because it needs to build a model for each type of action. This shows the rapidity of CNN-based method.

When the number of data samples increases to 9,600, the method in reference [23] takes longer, which is 61 times as long as the method proposed in this study, which fully proves the efficiency of the CNN-based method proposed in

FIGURE 9: Time-consuming comparison of upper and lower-limb movement experiments in data set I.

FIGURE 10: Time-consuming comparison of upper and lower-limb movement experiments in data set II.

this study. The recognition method based on CNN is three times faster than the recognition method in reference [24]. The reason is that the network structure in reference [24] contains self-connection, and the update calculation of the central unit is also time consuming. Therefore, compared with the existing methods, the proposed method based on CNN not only achieves the best recognition accuracy but also performs best in time performance.

5.2. Analysis of Upper-Limb Motion Segmentation and Labanotation Generation Algorithm. More than 400 upper-limb movement dances were collected in the experiment, including 232 left arm movements and 222 right arm movements, each with about 200 frames, for a total of tens of thousands of frames in the data set. In this experiment, 50% of the samples are used for training, and the remaining 50% of the data are used for testing, with the final result being the average of 10 experimental results.

The recognition accuracy of human upper-limb movements is first compared to a feature of normalized human orientation extracted from literature [28] using the same spatial segmentation method, as shown in Figure 11. Feature1 denotes the original action data, Feature2 the features added with normalization of human body orientation,

FIGURE 11: Comparison of different feature experiments.

FIGURE 12: Comparison of algorithm accuracy.

Feature3 the bone features of other nodes of the human body relative to the original nodes, and Feature4 the bone features of other nodes with normalization of human body orientation relative to the original nodes.

As can be seen from Figure 11, the bone features designed in this study have achieved the best recognition results, reaching about 94% recognition accuracy, which is about 2% higher than the features extracted from literature [28], which fully proves the effectiveness of the bone node vector features proposed in this study.

Figure 12 is the result of accuracy calculated by different dance movements. In the experiment, 50 pieces of dance data were collected, with 30,000 human body movement data of about 3 minutes per dance. Each piece of dance

action corresponds to 50 Laban symbols, including a series of actions such as walking, running, jumping, and arm rotation. It can be found that the spatial segmentation method is much higher than the other two algorithms in the action recognition of upper-limb nodes.

Through comparative analysis, we can see that the automatic generation algorithm of labanotation based on DL designed in this study can realize the efficient generation from motion capture data to labanotation and basically meet the needs of practical application.

6. Conclusion

In this study, the notation rules and fundamental elements of labanotation are thoroughly discussed, as well as the structural analysis and data conversion of BVH files. On this foundation, a motion capture data to labanotation conversion algorithm is proposed, and finally, automatic labanotation generation is realized. The DL-based segmentation algorithm and the spatial segmentation-based recognition algorithm were introduced for the first time, effectively improving recognition accuracy. CNN is primarily used to study and express the spatial characteristics of human body movements, and it can comprehensively learn the spatial position information of human body movements. To generate a Laban symbol sequence, the coordinates in the world coordinate system are projected into three-dimensional space, and the minimum beat unit is then incorporated into the Laban symbol sequence for regular optimization. Laban keeps track of the human body's movement in the simplest form possible. Finally, in order to obtain an accurate recognition result, this study introduces a classifier to eliminate the interference space in the middle. The segmentation and recognition are integrated directly from the spatial sequence in this study, which introduces a novel concept from data to dance music.

As labanotation is in the preliminary research stage in China, it still needs our constant exploration and research. At last, labanotation has become a universal communication tool in the international dance field, and it is hoped that Laban automatic generation system will be more and more perfect.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors do not have any possible conflicts of interest.

Acknowledgments

This study was supported by Research Start-up Fund for Young Teachers of Shenzhen University (project no. 000002111006) and Research Start-up Fund for Young Teachers of Shenzhen University (project no. 000002111105).

References

- [1] J. Wang, Z. Miao, X. Zhang, Z. Miao, and J. Wang, "Automatic generation of Labanotation for national dynamic art digitalization," *Scientia Sinica Informationis*, vol. 49, no. 2, pp. 229–244, 2019.
- [2] C. Cui, J. Li, D. Du, H. Wang, P. Tu, and T. Cao, "The method of dance movement segmentation and labanotation generation based on rhythm," *IEEE Access*, vol. 9, p. 1, 2021.
- [3] X. Zhang, Z. Miao, and Q. Zhang, "Skeleton-based automatic generation of Labanotation with neural networks," *Journal of Electronic Imaging*, vol. 28, no. 2, p. 1, 2019.
- [4] M. Li, Z. Miao, X.-P. Zhang, W. Xu, C. Ma, and N. Xie, "Rhythm-aware sequence-to-sequence learning for labanotation generation with gesture-sensitive graph convolutional encoding," *IEEE Transactions on Multimedia*, vol. 99, p. 1, 2021.
- [5] X.-B. Jin, W.-Z. Zheng, J.-L. Kong et al., "Deep-learning temporal predictor via bidirectional self-attentive encoder-decoder framework for IOT-based environmental sensing in intelligent greenhouse," *Agriculture*, vol. 11, no. 8, p. 802, 2021.
- [6] Y. Jiang, J. Zheng, X. Gu, J. Xue, and P. Qian, "A novel synthetic CT generation method using multitask maximum entropy clustering," *IEEE Access*, vol. 7, pp. 119644–119653, 2019.
- [7] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, Article ID 102048, 2021.
- [8] Z. Wu and W. Chu, "Sampling strategy analysis of machine learning models for energy consumption prediction," in *Proceedings of the 2021 IEEE 9th International Conference on Smart Energy Grid Engineering (SEGE)*, pp. 77–81, IEEE, Oshawa, Canada, August 2021.
- [9] Z. Li, X. Zhou, and Z. Ye, "A formalization model transformation approach on workflow automatic execution from CIM level to PIM level," *International Journal of Software Engineering and Knowledge Engineering*, vol. 29, no. 9, pp. 1179–1217, 2019.
- [10] J. Mariño, R. N. N. Alborodo, L.-A. Fredlund, and Á. Herranz, "Synthesis of verifiable concurrent Java components from formal models," *Software and Systems Modeling*, vol. 18, no. 1, pp. 71–105, 2019.
- [11] J. Park, G. Destgeer, M. Afzal, and H. J. Sung, "Acoustofluidic generation of droplets with tunable chemical concentrations," *Lab on a Chip*, vol. 20, no. 21, pp. 3922–3929, 2020.
- [12] J. Wang, Z. Miao, and N. Xie, "Labanotation generation from motion capture data for protection of folk dance," *IEEE Access*, vol. 99, p. 1, 2020.
- [13] P. Tobias, D. Lorenzo, and Z. Christoph, "Compator: a toolbox for the automatic generation of moment equations for dynamic compartment populations," *Bioinformatics*, vol. 17, p. 17, 2021.
- [14] L. Lin, T. Masopust, W. M. Wonham, and R. Su, "Automatic generation of optimal reductions of distributions," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 896–911, 2019.
- [15] A. A. Ylmaz, M. S. Guzel, and E. Bostanci, "A novel action recognition framework based on deep-learning and genetic algorithms," *IEEE Access*, vol. 8, no. 1, pp. 100631–100644, 2020.
- [16] P. Wang, "Research on sports training action recognition based on deep learning," *Scientific Programming*, vol. 2021, Article ID 3396878, 8 pages, 2021.
- [17] Y. Wan, Z. Yu, and Y. Wang, "Action recognition based on two-stream convolutional networks with long-short-term spatiotemporal features," *IEEE Access*, vol. 99, p. 1, 2020.
- [18] Y. Han, S.-L. Chung, Q. Xiao, W. Y. Lin, and S.-F. Su, "Global spatio-temporal attention for action recognition based on 3D human skeleton data," *IEEE Access*, vol. 8, pp. 88604–88616, 2020.
- [19] J. Zhu, W. Zou, Z. Zhu, and Y. Hu, "Convolutional relation network for skeleton-based action recognition," *Neuro-computing*, vol. 370, pp. 109–117, 2019.
- [20] C. N. Phyo, T. T. Zin, and P. Tin, "Deep learning for recognizing human activities using motions of skeletal joints," *IEEE Transactions on Consumer Electronics*, vol. 2, p. 1, 2019.
- [21] G. Xiao, "Research on athlete's action recognition based on acceleration sensor and deep learning," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2229–2240, 2021.
- [22] A. Muthukumar, "Smart phone ROM measurements based on action recognition in rehabilitation using deep learning," *International Journal of Advanced Science and Technology*, vol. 29, no. 12s, pp. 1257–1267, 2020.
- [23] Z. Jiao, G. Jia, and Y. Cai, "Ensuring computers understand manual operations in production: deep-learning-based action recognition in industrial workflows," *Applied Sciences*, vol. 10, no. 3, p. 966, 2020.
- [24] D. Wang, J. Yang, Y. Zhou, and Z. Zhou, "Human action recognition based on deep network and feature fusion," *Filomat*, vol. 34, no. 15, pp. 4967–4974, 2020.
- [25] Y. Jiang, J. Zhu, X. Gu et al., "Recognition of large-scale ncRNA data using a novel multitask cross-learning 0-order TSK fuzzy classifier," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 2, pp. 502–507, 2020.
- [26] J. Zhu, K. Li, K. Xia et al., "A novel double-index-constrained, multi-view, fuzzy-clustering algorithm and its application for detecting epilepsy electroencephalogram signals," *IEEE Access*, vol. 7, pp. 103823–103832, 2019.
- [27] L. Wang, J. Xue, C. Chen et al., "Urinary liver-type fatty acid-binding protein predicts recovery from acute kidney injury," *Clinical Nephrology*, vol. 84, no. 5, pp. 255–261, 2015.
- [28] S.-W. Sun, B.-Y. Liu, and P.-C. Chang, "Deep learning-based violin bowing action recognition," *Sensors*, vol. 20, no. 20, p. 5732, 2020.
- [29] J. Zhang, L. Zi, Y. Hou, M. Wang, W. Jiang, and D. Deng, "A deep learning-based approach to enable action recognition for construction equipment," *Advances in Civil Engineering*, vol. 2020, Article ID 8812928, 14 pages, 2020.
- [30] M. Al-Faris, J. Chiverton, D. Ndzi, and A. I. Ahmed, "A review on computer vision-based methods for human action recognition," *Journal of Imaging*, vol. 6, no. 6, p. 46, 2020.