Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 2, pp. 1073–1090, DOI 10.12694/scpe.v25i2.2648

SENSOR BASED DANCE COHERENT ACTION GENERATION MODEL USING DEEP LEARNING FRAMEWORK

HANZHEN JIANG *AND YINGDONG YAN †

Abstract. Dance Coherent Action Generation is a popular research task in recent years to generate movements and actions for computer-generated characters in a simulated environment. It is sometimes referred to as "Motion Synthesis". Motion synthesis algorithms are used to generate physically believable, visually compelling, and contextually appropriate movement using motion sensors. The Dance Coherent Action Generation Model (DCAM) is a generative framework for producing aesthetically pleasing movements using deep learning from small amounts of data. By learning an internal representation of motion dynamics, DCAM can synthesize long sequences of movements in which coherent patterns can be created through latent space interpolation. This framework provides a mechanism for varying the amplitude of the generated motion, allowing further realistic thinking and expression. The proposed model obtained 93.79% accuracy, 93.79% precision, 97.75% recall and 92.92% F1 score. DCAM exploits the balance between imitation and creativity by enabling the production of novel outputs from limited input data and can be trained in an unsupervised manner or fine-tuned with sparse supervision. Furthermore, the framework is easily extended to handle multiple layers of abstraction and can be further personalized to a particular type of movement, enabling the generation of highly individualized outputs.

Key words: Dance, Coherent, Action, Generation, Computer Animation, Video Games,Sensor.

1. Introduction. Dance is an activity that allows people to express themselves and cultivate relationships through creative and meaningful body movements. Recently, it has developed into a popular form of art and recreation. It is a great way to promote physical activity and socialization and build self-confidence. Generating coherent action in dancing is a process that involves developing knowledge of the various techniques, steps, and styles associated with a chosen dance [1]. It can be done by studying the history and culture behind a particular dance style, understanding its body movements, and learning the appropriate techniques and steps through practice. In order to generate coherent action in dancing requires analytical thinking and creative expression. It involves understanding the context in which the dance takes place, such as the atmosphere, music, and the audience. It also involves being able to recognize the dynamics and nuances of the dance and have the ability to be flexible in order to adapt to the changing environment [2]. It is essential to have an idea of the structure of the dance and how it can be manipulated to create the desired choreography. Generating coherent action in dancing also necessitates body awareness and control. It is essential to identify the correct muscles and movements used in the dance and utilize them in a planned and strategic way [3]. Finally, it involves understanding how to convey emotion and rhythm through movement and developing a good sense of coordination, timing, and balance. Generating coherent action in dancing is an intricate process that requires analytical thinking, creativity, body awareness, and control. Through practice and dedication, one can successfully generate coherent action in dancing and bring to life a unique and enjoyable experience for all. Dance has existed since the beginning of time and is constantly evolving and innovating. In modern times, the development of technology has had a significant impact on how dance is created, performed, and taught [4]. One of the most significant innovations in dance is the development of coherent action generation. This technology allows skilled dancers to generate complex movements based on a given set of goals or objectives. By breaking down sophisticated choreographic processes into discrete elements, these movements can be performed and perfected much more quickly. It has allowed professional dancers to create stunning and complex routines without spending hours creating them from scratch. Another development within the dance industry is the use of motion capture technology [5]. This technology is used to map the movements of a dancer in order to create a 3-D interpretation of their routine.

*International College, Krirk University, Bangkok, 10220, Thailand (hanzhen93@sina.com)
†International College, Krirk University, Bangkok, 10220, Thailand (791086531@qq.com)

It allows the choreographer to analyze and adjust the routine as needed to perfect the performance. Thanks to motion capture technology, 3-D-created routines can also be studied and replicated by others. The technology used to create virtual simulations is advancement in dance [6]. With virtual simulations, dancers can visualize different stages of their performance in a manner that were not possible before. It is allowing 3-D choreography to become commonplace, even with opener-level dancers. This technology is used to plan out routines and in the editing process once a dance has been completed. It is essential to note the increased use of smart phones and other mobile devices in dance. These devices allow dancers to access instructional videos, lessons, and other resources from almost anywhere [7]. It has also allowed them to connect with choreographers and other dancers to further their skills. The advancements in technology are revolutionizing how dancers approach and perform their art. Through coherent action generation, motion capture, virtual simulations, and mobile devices, dancers can create complex routines and take their craft to new heights.

The research and practice of dance coherent action generation is a field of study that investigates how computers can create sequences of movements to provide a dancer or a group of dancers with a logical, fluid, and aesthetically pleasing pattern of movements. Deep Learning (DL) is crucial in this research as neural networks can learn from example data and generate predictive models [8]. This research and application of DL methodologies has become increasingly popular in dance and the automated generation of coherent motion sequences. One of the most successful approaches to dance coherent action generation is using Long Short-Term Memory (LSTM) networks. LSTMs can capture short-term and long-term dependencies in the motion of a dancer. Such networks can generate coherent motion sequences based on an initial input path and influential parameters, bypassing the traditional choreography process. In addition, such networks allow a controllable degree of randomness, which affects the generated motions, enhancing the expressive qualities of the dance sequence [9]. Combined with further recurrent neural networks, this model enables a projective technique that limits the rate of change and preserves the timing characteristics of sequences. DL networks have been successfully applied to various domains of the arts, such as music and visual arts. However, issues of aesthetics and expression have yet to be considerably explored in research on computer choreography. Recent Reinforcement Learning (RL) developments have provided promising solutions for this aspect of the field. Such solutions use a reward function and principles of imitation learning to generate coherent and expressive movements [10]. As computer choreography and DL-aided motion generation mature, these technologies will become more common as they present an opportunity to reduce the time and effort necessary for the manual creation of choreography. Nevertheless, further research is needed to address some of the limitations of current DL techniques for dance and accompanying expressive qualities, such as the ability to handle highly dynamic and expressively charged dances. The main contribution of the research has the following,

- Developing and validating metrics for evaluating the quality of dance motions and generating realistic motions using data-driven methods such as motion capture data;
- Exploring different ways of combining multiple motion sources to produce more fluid and consistent motions and developing motion models capable of learning to generate novel dance motion;
- Understanding the underlying formal grammar of dance motion and modeling physical dynamics for motion synthesis;
- To design efficient motion representation and scheduling algorithms for efficient online motion generation and developing interactive systems for generating coherent dance sequences.

2. Related Work. Muthukumaran, V et al. [11] have discussed Motion generation based on a multi-feature fusion strategy is a technique used to generate motion based on multiple features of a body or object. It requires the fusion of pre-experimentally acquired features from various sensors, such as motion sensors, wearable biosensors, and inertial measurement units. This technique is used to generate realistic and robust motions for robotic applications and animation and gaming applications. It is also used in medical applications, rehabilitation, and training applications in sports and other disciplines.

Zhang et al. [12] have discussed Mining and applying composition knowledge of dance moves for style-concentrated dance generation is a process that utilizes artificial intelligence techniques (such as machine learning and deep learning) to analyze existing dance moves and styles from different choreographers and create new dance sequences. The techniques used in this process can recognize patterns in existing choreography and apply the found knowledge to create unique dance sequences that reflect a particular style. As examples, it could include break dancing, ballet, or hip-hop. This process could revolutionize the field of dance composition, allowing dancers to create expressive and original dances much faster than traditional methods.

Aberman et al. [13] have discussed Deep video-based performance cloning as a technique used to capture and reproduce human performance, such as facial expressions, gestures, and movements. It uses deep learning algorithms to synthesize a clone of a person through videos or images. This technique can recreate an individual's entire performance from a video, replicating their exact facial expression, gestures, and movements. Deep video-based performance cloning has many applications, from virtual coaching to video game animation.

Zheng et al. [14] have discussed Time series data prediction and feature analysis of sports dance movements based on machine learning is a technology that can improve the accuracy, performance, and efficiency of predicting movements. It uses data that has been collected over some time in order to detect patterns and make predictions about future events. It can be used to predict everything from the action of an athlete when performing a specific move to the team's motion during a break in play. Using machine learning techniques can give coaches, players, and fans a better understanding of a sport's dynamics and help improve athletes' performance. It can also provide valuable insights into the pattern of movements associated with certain sports. By analyzing time series data, machine learning can also identify any discrepancies in an athlete's motion, such as variations in technique or forces. This technology can help coaches develop better training methods for their athletes and give players and fans a better understanding of the game they are playing.

Dong et al. [15] have discussed to improve the interpretability of deep neural networks with semantic information as a process of utilizing semantic information to enhance the behavior of deep neural networks. Deep neural networks are incredibly complex models, and understanding how they decide is difficult. Semantic information provides a way to make the decision-making process of a deep neural network more interpretable. Semantic information can add context to a model, helping explain why a specific decision was made. Semantic information can also reduce the complexity of explanations by categorizing results into more meaningful categories. This increased interpretability of deep neural networks ultimately helps to improve their performance and trustworthiness.

Sünderhauf et al. [16] have discussed the potential of deep learning for robotics is limitless. It can help in motion planning, visual perception, navigation, and decision-making. Deep learning algorithms can assist in advanced obstacle avoidance, automated driving, and home automation. The limits to deep learning in robotics are cost, infrastructure, and access to data. Cost is a primary limiting factor since deep learning algorithms require a lot of computing power to work, which can drive up the cost. Additionally, deep learning requires data to train appropriately, which can take time to acquire, depending on the industry. Finally, access to training and infrastructure can be a limitation as it can require specialized personnel with the right expertise.

Wu et al. [17] have discussed Image Comes Dancing With Collaborative Parsing-Flow Video Synthesis is a technique for creating videos from a single image. The technique involves using computer-based synthesis algorithms to analyze the source image, interpret it according to specific parameters, and produce a resulting video based on the input image. This technique has been used to create videos of dancing characters from a single image, with the characters dynamically dancing to match the rhythm of the music. It has also been used in various interactive video games and video-editing applications.

Kico et al. [18] have discussed Digitization and visualization of folk dances in cultural heritage is the process of digitally capturing, preserving and sharing the stories and movements of a variety of folk dances around the world. It can be achieved by recording and documenting folk performances, utilizing digital media, and creating interactive 3D computer visuals and animations. Digitization and visualization also help preserve and share the cultural heritage of folk dances for future generations, allowing for new interpretations of traditional motions. Ultimately, it serves as a way to spread cultural appreciation and create a bridge between different generations.

S. T. Ahmed et al. [19] have discussed dance style transfer with a cross-modal transformer is an artificial intelligence that takes insights from a source of dance-related material, such as online videos, and translates them into a different dance style. It does this by leveraging machine learning and natural language processing to try and capture the "essence" of a particular dance style. This technology has allowed for the rise of creative AI projects that can translate music, movie dances, or even user-created dance videos into different dance styles.

Cai et al. [20] have discussed An Automatic Music-Driven Folk Dance movement generation Method Based on a Sequence-To-Sequence Network is an AI-based approach to automatically generate folk dance movement sequences by predicting and following musical input. This method uses a deep learning model incorporating music and motion data to generate more realistic and creative folk dance movements according to the musical beat. This method utilizes a sequence-to-sequence network architecture, which allows for creating new dance sequences by using various input sources. This method is beneficial for quickly creating and customizing folk dance performances with different musical genres. It can also provide a more reliable and entertaining way to generate custom dances than traditional choreographers.

Li et al. [21] have discussed Human motion recognition in dance video images based on attitude estimation is a technique used in computer vision and visual recognition to identify the body posture of a dancer in a video image. This technique uses a feature-based attitudinal translation architecture to detect and classify the body poses of a dancer from a given dance video sequence. The primary goal of human motion recognition in dance video images based on attitude estimation is to accurately recognize the body poses of a dancer in an unstructured dance environment to facilitate automated motion analysis and evaluation of dance performances.

2.1. Research Gap. In this model, the number of low power and low cost sensor has been deployment as self-adaptive and traffic dependent network protocol on the traffic of the network. The node data transmissions of the sensed data are adaptively changes to the traffic pattern. Power changes occur based on traffic [9]. The node will be time- synchronized for path negotiation and data contention on basis of node density. Path Allocation model of the protocol enhances the transmission capabilities on the less utilized nodes to prevent network degradation. Further linear programming architecture has been employed to Dynamic node hopping sequence. The routing architecture provides optimal stability among the node transmission time with respect to node availability and energy consumed on the effective path within specified delay along throughput constraints to solve energy whole problem

- Limited research on combining motion capture with deep learning techniques for generating the motion of characters derived from dance actions and more work must be done on implementing choreographic structures within coherent action generation.
- Few studies have examined the use of motion capture and behavior recognition techniques to monitor and evaluate the effectiveness of a generated action sequence in a human-robot interaction.
- Few studies have explored the effects of different music styles on dance-style action generation and little work needs to be done on adapting generated motion to context and environmental factors.
- The use of AI to enable autonomous, cooperative, and emergent motion patterns is still in its infancy and research is needed to develop machine learning models that capture temporal dynamics better and create realistic and natural-looking motion.

2.2. Research Objectives.

- To explore and generate novel dance movements and sequences through advanced deep learning architectures and strategies for dance motions.
- To develop an AI-based model, that is capable of creating movement variations to generate unique and appealing dance choreography
- To investigate the affective context of the dance motion by training different deep learning networks and gains an understanding of the underlying principles of motion in human-like dancing styles.
- To discover the different factors that influence dancing/motion in different rhythms and beats and design appropriate strategies for automated dance generation with respect to small details of the movements.
- To explore methods to improve the fluency and accuracy of predicted dance motions and construct an effective and efficient training environment for deep learning models to generate dancing sequences.

3. Methodology. The proposed dance coherent action generation model based on deep learning is a model for the automated generation of dance moves from videos. It applies deep learning to extract motion features from videos that represent the movement of dancers and use these features to generate new dance moves. The dance coherent action generation system has shown in the following fig. 3.1. The model is designed to generate smooth transitions between different moves and generate new moves that align with the video's overall style. It can also control the speed of the generated dance moves, creating high-quality visualizations of dancing videos. The model is trained on various dance data sets, which helps provide more accurate results.

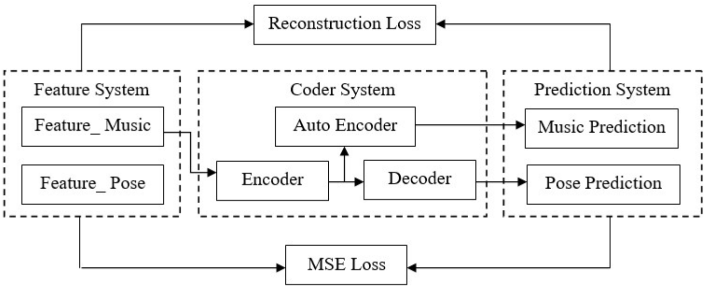

Fig. 3.1: Dance coherent action generation system

3.1. Feature system. It is a dance coherent action generation model extracts meaningful features from the environment. These features are used to inform the model of the state of the environment and to guide the decision-making process for artificial agents. These features can range from visual cues such as colors, shapes, and textures to auditory cues such as music tempo and pitch and kinetic features such as movements. The system also allows feature extraction from multiple domains, such as vision, audio, and motion. It allows for more robust and adaptive decision-making for the artificial agents in the model. Furthermore, feature extraction can detect global or local environmental patterns, making it easier for the model to generate dynamic and fluently choreographed dance sequences.

3.2. Coder system. It is an essential to the dance coherent action generation model. It performs a vital role in generating all the actions that constitute a dance. Its primary function is to encode the motion and music information into a format the system can understand. It does this by extracting meaningful visual and audio data from the dance video and associating it with different labels representing the motions and notes. The coder system then sends this encoded data to the agent, which then uses it to generate the sequences of actions that will constitute the dance. It ensures that the generated action will be coherent and consistent without discrepancies. The coder system also allows the model to “tune itself” by adjusting the resolution of the actions in order to reflect the desired motion better.

3.3. Prediction systems. It has dance-coherent action generation models are responsible for predicting the upcoming action of a dancer in order to generate a coherent dance routine. These systems are used to generate an output, which is a sequence of movements that correspond to a specific dance style, and it aims to produce a more realistic performance compared to pre-programmed dance steps. Prediction systems are used to interpret and anticipate the body movements of a dancer and generate appropriate sequences of movements in response. They use deep learning, statistical analysis, and motion capture techniques to track the dancer’s movement. Furthermore, they use the previous movements to anticipate the next movement, making it easier for the model to generate a coherent and connected dance routine for the dancer.

3.4. Reconstruction loss. It is a function used in the Dance Coherent Action Generation Model. It is used to optimize the model according to the human motion data that is being used. Reconstruction loss works by computing the difference between the model’s output and human motion data used for the input. The model adjusts its outputs by weighting and reducing the error resulting from the difference between the expected output and the actual human motion data. It means the model is more accurate, allowing it to predict human motions better. Reconstruction loss allows the Dance Coherent Action Generation Model to predict human motion better, providing more accurate output for the user.

3.5. MSE (Mean Squared Error) loss. It is a widespread loss function frequently used in machine learning models. In a dance coherent action generation model, this loss function calculates the errors between the predicted result and the real action. It measures the average of the squares of the errors or deviations, which is then used to gauge the model’s efficiency and improve its accuracy over time. This loss function is commonly used in supervised learning tasks like the dance coherent action generation model. By taking the

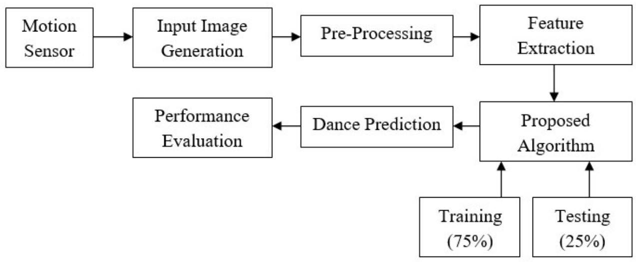

Fig. 4.1: Proposed block diagram

differences in prediction as errors, MSE loss can lead to improved predictions and more accurate generation of dance coherent actions.

A “coherent action” is a behavior that results in a composition of physical movements that is organized, stress-free, and capable of producing unified motion. It is a physical expression that is purposeful and organized, connecting body parts in an intrinsic manner that is relatively free from perceived tension. This definition specifically applies to the field of dance, where movement not only expresses emotion but also melds space, time, and music into a storytelling experience. In the field of dance, sensors, and sensor data allow information to be collected on how a dancer moves through space and can be used to create and execute more precise or customized actions for a choreographed routine. Sensors that can be used range from physical to digital and are used to create an interactive performance environment. The most commonly used physical sensors are motion-capture systems, accelerometers, inclinometers, and gyroscopic sensors. Digital sensors include cameras, microphones, and microphone arrays. These sensors and sensor data are used to generate a visualization of a dancer’s body, creating a visualization of the dancer’s movement, body shape, gesture, and interaction. This data is then analyzed and evaluated to produce an action that is meaningful and articulate. Ultimately, with the assistance of sensors and sensor data, a dancer’s routine can be designed, and the body position of the dancer can be monitored, resulting in a more precise action and a more accurate representation of the idea being expressed.

4. Proposed Model. The Dance Coherent Action Generation Model Using Deep Learning generates realistic, engaging, and fluid movement using Deep Neural Networks. It allows for a wide range of motion styles, from traditional hip-hop and freestyle movement to contemporary styles like jazz and Lyrical. The proposed block diagram has shown in the following fig. 4.1.

The model works by taking in raw video or motion-capture data of a dancer or performer’s body movements and producing motion-capture data that can be used for animation. It can fuse styles seamlessly, allowing for a much more comprehensive range of motion styles than traditional animation tools. The model can also divide a single movement style into multiple parts, allowing for speed and elegance. Finally, it can create new motion combinations from existing movement data to help innovate traditional dance styles. The proposed framework is shown in Table 4.1.

The proposed model can be applied to generate different types of dance styles by using specific data sets, such as motion capture data or motion synthesis, depending on the type of dance. This model can also be used for generating natural motion sequences, such as those seen in expressive dance. The proposed model should be evaluated against other existing models to analyze its efficacy in action generation. Existing action generation models can be used for comparison. It is also important to compare the qualities of the generated motion sequence with those of actual dance styles. The model should be evaluated in regards to its ability to be adapted to different body types and dancing abilities. Different body types and motion preferences can be encoded in the model to generate meaningful and personalized dance movements. Additionally, using different kinematic data from different dance skills can help refine the generated motion sequences to match desired levels of difficulty. Scalability is an important factor to consider when developing models. When expanding a dataset or model, one typically wants to do so in a manner that is as efficient as possible. The following are

Table 4.1: Dance Coherent Action Generation framework

1.	IP: a //Initial dance action;
2.	OP: b //Final dance action;
3.	Start
4.	GET_Sensor_IP ( ); //Get the sensor input data;
5.	Compute the a (t) ;
6.	if t=1
7.	PoP (t) = Compute the P ( ); // Computation of preliminary population;
8.	While (stop_con( )) do
9.	EVA_ the PoP (t); // Evaluation;
10.	CAP_ IDA; // Capture the initial dance action;
11.	ADJ_F(a); // Adjust the fitness function;
12.	SEL_Next ( ); // Select the next action;
13.	GEN_PoP$_{re}$ ( ); //Generate repeat dance movements;
14.	GEN_PoP$_{co}$ ( ); //Generate cross over movements;
15.	GEN_PoP$_{mc}$ ( ); //Generate motion capture function;
16.	PoP (t+1) = (PoP$_{re}$)+(PoP$_{co}$)+(PoP$_{mc}$);
17.	t = t+1; //Incremental operation;
18.	End;

some basic strategies that can be used to ensure scalability:

Automated solution: Automated solutions such as auto-scalers and autoencoders can be used to enable the model to scale up or down as the dataset grows or shrinks.

Parallelization: Parallelizing the dataset or model allows for the completion of tasks in much shorter timeframes while keeping scalability efforts to a minimum.

Establish a baseline: Establishing a baseline provides an organized structure upon which your model can be built. It also gives you a place to compare and identify areas where you can improve your scalability efforts.

Preprocessing: Preprocessing data can help simplify a dataset, making it easier to interpret and use for predictions, which can help the model become more scalable.

Feature Selection: Automated feature selection can help improve accuracy and reduce computational complexity while maintaining scalability.

Model Evaluations: Regular evaluations of the model to assess scalability can help determine what strategies need to be implemented (or improved) in order to achieve desired scalability.

Observed sequences of actions: Leveraging observed sequences of actions can provide predictive models that can be used to identify valuable insights from the data. It can allow for more informed decisions to be made and is an important tool for scalability.

The model is not necessarily robust enough to generate coherent actions in various environments with different sensor configurations. Different environments and sensor configurations might require different approaches to agent behavior. Further fine-tuning of the model would be necessary to make it robust enough for various environments.

4.1. Pre-processing. It is a significant step in proposed model and is especially beneficial for a dance-coherent action generation model. Pre-processing can help to normalize the input data, clean up noise, and reduce redundancies. Most importantly, it can reduce the dimensionality of the input data, making it easier for the algorithm to learn a profound representation of the data. The upper bound dance actions has indicated in the following eq. 4.1

$$\forall a > 0; \forall a \in X \Rightarrow x \leq feature\_X; \tag{4.1}$$

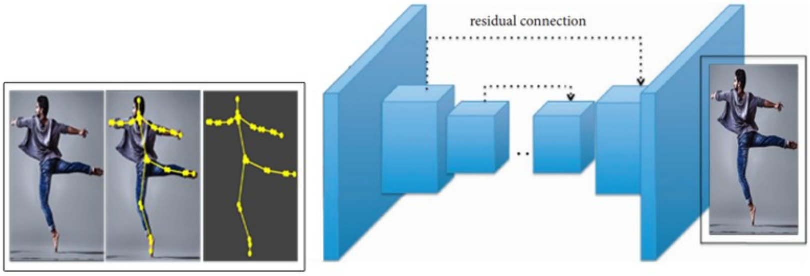

Fig. 4.2: Dance Image Generator

where a is an initial dance action, X is a upper bound dance segment and x is a coherent function. The lower bound dance actions has indicated in the following eq. 4.2

$$\forall b > 0; \forall b \in Y \Rightarrow y \leq feature\_Y; \tag{4.2}$$

where b is an final dance action, Y is a lower bound dance segment and y is a coherent function. The convexity function has shown in the following,

$$UB : q(\alpha s_1 + (1-\alpha)s_2) \leq \alpha q(s_1) + (1-\alpha)q(s_2) \tag{4.3}$$

$$LB : q(\alpha s_1 + (1-\alpha)s_2) \geq \alpha q(s_1) + (1-\alpha)q(s_2) \tag{4.4}$$

where UB shows the upper bound dance action and LB shows the lower bound dance action. The Dance Image Generation has shown in the following fig. 4.2.

Pre-processing can also help fill any gaps in the data or scale the features so that different data sources can be used together. Finally, it can reduce the computational complexity associated with complex deep neural networks, allowing the model to more quickly and effectively. The sigmoid function has shown in the following eq. 4.5

$$A(z) = \left\{ \frac{1}{1+e^{-z}} \right\} \tag{4.5}$$

The image generator in the dance coherent action generation model using deep learning is a convolutional neural network-based architecture. This model uses a long short-term memory (LSTM) layer as a guide for generating image frames, which are then given as inputs to the convolutional neural network. The directional derivatives has shown in the following,

$$\frac{\partial q}{\partial d} = \frac{\partial q}{\partial s}\cos\phi + \frac{\partial q}{\partial c}\sin\phi \tag{4.6}$$

The convolutional neural network then passes the image frames through convolutional layers, which extract features. These features are then used to generate the next image frame. This process is repeated until the entirety of the motion is generated. The model then produces a sequence of motion vector representations used to reconstruct the motion as a video clip or animation. The linear interaction has shown in the following eq. 4.7

$$a_x = \beta(C_e * [d_{x-1}, f_x] + g_e) \tag{4.7}$$

$$b_y = \beta(C_e * [d_{y-1}, f_y] + g_e) \tag{4.8}$$

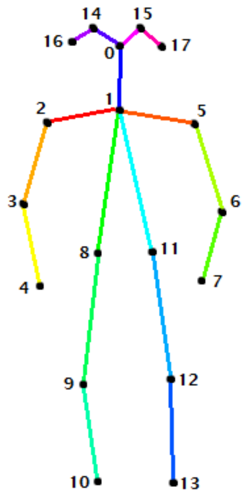

Fig. 4.3: Human body feature points

where a and b are the dance action, β is the coherence factor, x and y are the motion sensor coordinates. C is the coder function. The image generator in a deep learning model for dance coherent action generation visualizes the input movements. This visual representation is then used to generate a smooth, coherent sequence of output movements. It serves as an efficient way to efficiently produce an action from the input motion, thus preserving the continuity of the motion. Additionally, the image generator is used to create a visual representation of the output sequence used as a reference to compare the motion generated by the model.

4.2. Feature Extraction. Feature extraction in deep learning is extracting meaningful features from data. It is a crucial step in any deep learning model, as it helps to identify patterns in the data and determine which features are essential to capture. In proposed model, feature extraction is used to create valuable representations of motion sequences, which can then be used as input for the model.

$$a_x = \beta(C_e * d) \tag{4.9}$$

$$b_y = \beta(C_e' * d') \tag{4.10}$$

$$C' = C^T \tag{4.11}$$

$$L(ab) = \|a - b\|^2 \tag{4.12}$$

$$L_X(a,b) = -\sum_{m=1}^{n} \{a_m \log b_m + (1 - a_m) \log(1 - b_m)\} \tag{4.13}$$

where a is the initial dance action, b is the final dance action, C represent the coder function. It allows the model to identify better and generate the correct dance moves. By using feature extraction on motion sequences, the model can better differentiate between features important to replicating a dance move and those not. Fig. 4.3 shows the human body feature points.

The human body feature points in the dance coherent action generation model refer to points on the body detected by a camera and used to predict movements and generate performers' dance motions. These points are typically used to capture motion and are placed on essential body parts, such as the arms, legs, torso, and head joints. Table 4.2 shows the human body feature points in detail.

Table 4.2: Human body feature points

S.No	Feature Point	Representation
0	Dancer Nose	a1, a2
1	Dancer Neck	b1, b2
2	Dancer Shoulder (Right)	c1, c2
3	Dancer Elbow (Right)	d1, d2
4	Dancer Hand (Right)	e1, e2
5	Dancer Shoulder (Left)	f1, f2
6	Dancer Elbow (Left)	g1, g2
7	Dancer Hand (Left)	h1, h2
8	Dancer Leg (Right)	i1, i2
9	Dancer Knee (Right)	j1, j2
10	Dancer Foot (Right)	k1, k2
11	Dancer Leg (Left)	l1, l2
12	Dancer Knee (Left)	m1, m2
13	Dancer Foot (Left)	n1, n2
14	Dancer Eye (Right)	o1, o2
15	Dancer Eye (Left)	p1, p2
16	Dancer Ear (Right)	q1, q2
17	Dancer Ear (Left)	r1, r2

This information is then used to interpolate between poses and motions to generate a realistic and natural-looking output. For example, these feature points can represent a dancer's motion, such as twisting, turning, and stepping, and can generate a realistic and synchronized version of the dancer's motion. This way, the model can generate a natural and accurate representation of the dancer's motions. The improved training process in the Dance Coherent Action Generation Model using Deep Learning involves a two-stage approach. In the first stage, the model is trained on synthesized sequences of postures adapted from go-go dance videos. It allows the model to learn the general motion structure of the dance. In the second stage, the model is then fine-tuned using real-world data. This fine-tuning allows the model to model the nuances of real-world dancing and efficiently produce realistic results. The improved training process has shown in the following fig. 4.4.

As a result, the improved training process provides a more accurate representation of the physical body movements of go-go dancing. It ultimately helps in creating more realistic and coherent dance sequences.

4.3. Detection. Detection in dance coherent action generation models using deep learning allows the model to detect specific features in the input image or video sequence data used to generate dance steps autonomously. In the context of neural network models, detection typically refers to providing a model with the necessary visual cues required to generate the desired output. The detection of motions using the sensor has the following,

$$a_m = \sum_{x=1}^{T} \{\delta_{x,m} * g_x\} \tag{4.14}$$

Where, a is the input dance detection in mth state, δ is the attention weight, x and m are the inputs of deep learning model. These visual cues can be in the position, size, or orientation of particular objects in the image. In a dance setting, the model must detect the dancer's body position, movement patterns, and the context of the dance scene to generate coherent movements. To obtain the detection of dimensional dance moves has the following,

$$L^x = \mu \left\{C, L^{x-1}\right\} \tag{4.15}$$

$$R^{T_x} = \omega^x \left\{C, L^{T_x}\right\} \tag{4.16}$$

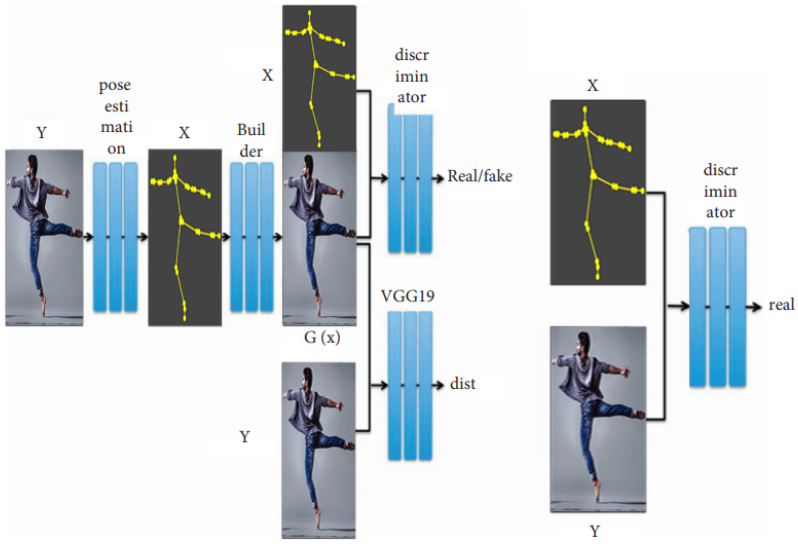

Fig. 4.4: Improved training process

$$R^T = \omega^x \left\{C, L^{T_x}, R^{x-1}\right\} \tag{4.17}$$

$$F = \min_a \max_b L_x(a,b) + \alpha_x L_{MSE}(a) + \alpha_y L_x(E,D) \tag{4.18}$$

where F is the final detection of dance coherent action, a is the initial dance action and b is the final dance action. MSE shows the mean square error, E is the encoder and D is the decoder. The proposed model may be used to detect the location and movement of the dancer's limbs, which can then be used to generate the best set of moves that synchronize with the music. Detection is necessary for deep learning models as it allows them to acquire visual cues to generate the desired output.

4.4. Classification. Classification in dancing is used to categorize the types of movements a dancer performs [22]. Regarding consistent action generation using deep learning, classification allows the model to identify and recognize patterns among dance moves and use that knowledge to generate new, similar moves. Dance Coherent Analysis is an important part of the Coherent Action Generation Model using Deep Learning. This model uses deep learning to generate realistic motion of a dancer that is both expressive and organic, while still being physically and aesthetically pleasing. The Dance Coherent Analysis module consists of an analytical pipeline to detect and analyze the motion data of the dancer. It extracts features related to body positions and movements from the footage, estimates the dancer's plans, and extracts motion metrics that can be used to describe the performance of the dancer. The extracted features are then used to generate a trained behavior model for a dancer. The resulting model is used to animate a dancer in a simulated world, to generate dance-based applications. This model helps animate realistic dancing motions with few parameters and makes it much easier to generate realistic animations. From eq.18, the Lx(a,b) and LMSE(a) has classified as the following,

$$L_x(a,b) = F_{(c,d)}\left\{\log b(c,d)\right\} + F_c\left\{\log(1 - b(c, a(c)))\right\} \tag{4.19}$$

$$L_{MSE}(a) = F_{(c,d)}\left\{\log b(c_{e-1}, c_e, d_{e-1}, d_e)\right\} + F_c\left\{\log(1 - b(c_{e-1}, c_e, a(c_{e-1}), a(c_e)))\right\} \tag{4.20}$$

The dance coherent analysis has shown in the following fig. 4.5.

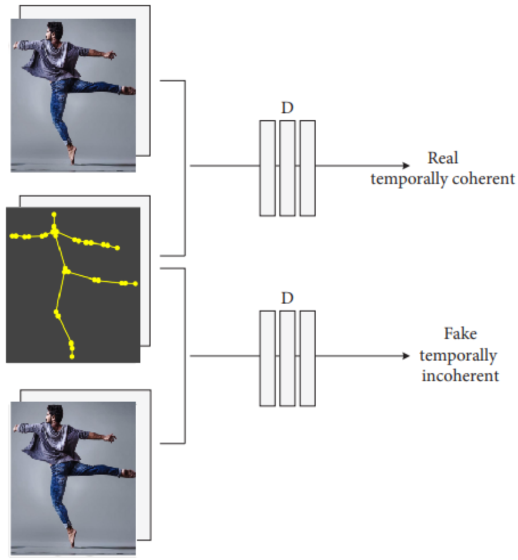

Fig. 4.5: Dance coherent analysis

Classification also helps the model recognize and understand the elements of the dance moves to generate new steps that fit within the constraints of the choreography. It allows the model to produce smooth, realistic dancing sequences that feel natural and coherent.

Sensor-based dance coherent action generation model using a deep learning framework is an innovative approach to using a combination of sensing, processing, and predictive algorithms to generate ground-breaking, impactful, and expressive human-like dances. By combining sensing and deep learning techniques, this model allows for optimized and automated processes of sensor data, which can enable truly autonomous generated dances. The model employs algorithms to learn the trends in motion performance from dance sequences, the variations in movements, and the correlations between different movements. It provides a potential performance outcome on the basis of the captured motion data in the form of a probability distribution instead of a single prediction. This approach improves the motion performance, fosters an emotionally expressive performance, and increases efficiency while reducing user input. This model is also scalable to multi-sensor dance training, making it a powerful tool for creating larger-scale, intelligent, and expressive dances through machine learning.

Learning systems that are able to detect relevant information from sensors and generate coherent actions accurately are typically based on artificial neural network algorithms. These systems use the input from the sensors to learn the characteristics associated with a particular type of behavior. Once the system has enough data, it can be trained to recognize similar patterns and generate actions accordingly. For example, an object recognition system can identify different objects in an image and process this information so that it can take an appropriate action. Additionally, reinforcement learning algorithms can be used to teach an agent to find the optimal path to a given goal based on various types of feedback from the environment. With the help of supervised learning algorithms, the system can learn to recognize specific patterns and generate appropriate responses.

5. Comparative Analysis. The proposed dance coherent action generation model (DCAGM) has compared with the existing Recurrent Neural Network (RNN), Dance Based on Deep Learning (DBDL), Quantum-based creative generation method (GCGM) and Convolutional sequence generation (CSG). Here python is a simulation tool used to execute the results.

5.1. Computation of Accuracy. Accuracy refers to the percentage of correctly classified examples in a deep learning dance coherent action generation model. The number of correctly classified examples must be

Table 5.1: Comparison of Accuracy (in %)

No.of Inputs	RNN	DBDL	GCGM	CSG	DCAGM
100	65.05	58.72	72.07	74.22	97.46
200	63.56	56.75	69.65	72.02	95.47
300	62.76	55.62	69.24	71.22	94.27
400	60.43	54.41	67.64	70.55	93.79
500	59.42	54.04	65.32	69.12	92.36
600	58.78	52.51	64.07	68.03	92.20
700	58.12	52.01	61.34	67.55	91.43

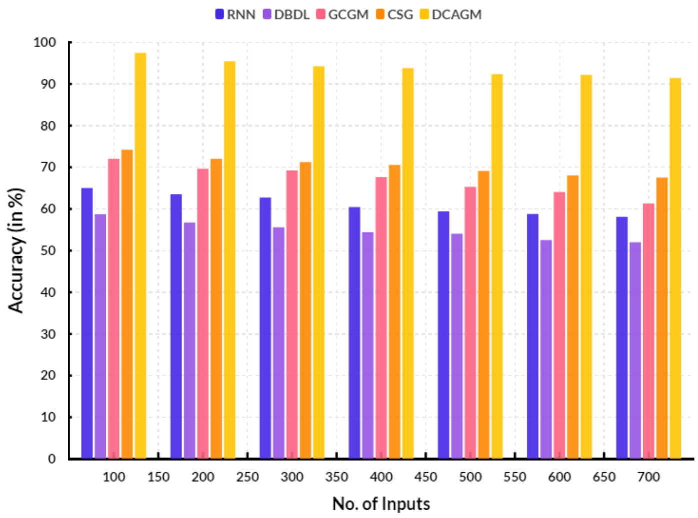

Fig. 5.1: Accuracy

divided by the number of examples examined. The comparison of accuracy has shown in the following table 5.1. It is a widespread loss function frequently used in machine learning models. In a dance coherent action generation model, this loss function calculates the errors between the predicted result and the real action. It measures the average of the squares of the errors or deviations, which is then used to gauge the model's efficiency and improve its accuracy over time. This loss function is commonly used in supervised learning tasks like the dance coherent action generation model. By taking the differences in prediction as errors, MSE loss can lead to improved predictions and more accurate generation of dance coherent actions.

Fig. 5.1 shows the comparison of accuracy. In a computation tip, existing RNN obtained 60.43%, DBDL obtained 54.41%, GCGM obtained 67.64%, and CSG obtained 70.55% accuracy. The proposed DCAGM obtained 93.79% accuracy. The proposed model does make use of transfer learning techniques to improve the accuracy of action generation. Transfer learning involves utilizing information and knowledge gained from existing models to improve the performance of new models. Transfer learning can be used to take data from previously trained models and apply it to new tasks, helping the model learn without having to start from scratch. In the proposed model, the transfer learning technique is used to help the model learn more quickly and accurately from previous data to improve its performance when generating new actions.

5.2. Computation of Precision. The computation of precision has to predict the action sequences and the actual test data used to train the model. This task is usually done by calculating the overall accuracy, which is the proportion of correctly identified correct actions versus the total number of observations in the test

Table 5.2: Comparison of precision (in %)

No.of Inputs	RNN	DBDL	GCGM	CSG	DCAGM
100	75.05	68.72	82.07	84.22	97.46
200	73.56	66.75	79.65	82.02	95.47
300	72.76	65.62	79.24	81.22	94.27
400	70.43	64.41	77.64	80.55	93.79
500	69.42	64.04	75.32	79.12	92.36
600	68.78	62.51	74.07	78.03	92.20
700	68.12	62.01	71.34	77.55	90.43

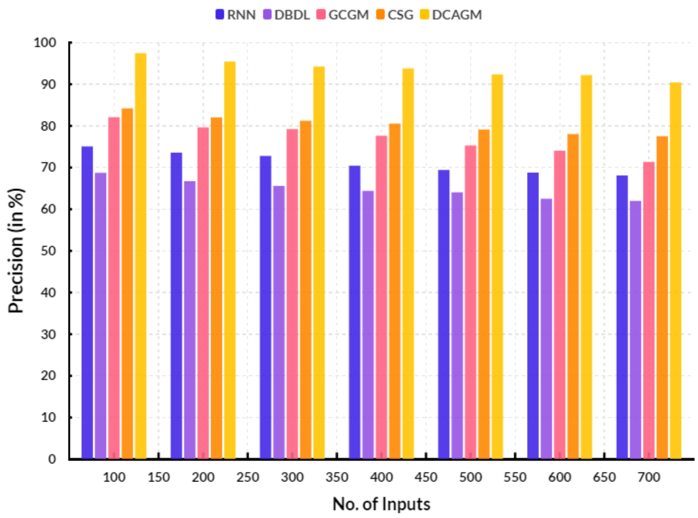

Fig. 5.2: Precision

data. In addition to the overall accuracy, precision of the predictions is also measured. Precision reflects the ability of the model to identify the correct dance actions with positive examples only correctly. The comparison of precision has shown in the following table 5.2.

Fig. 5.2 shows the comparison of precision. In a computation tip, existing RNN obtained 70.43%, DBDL obtained 64.41%, GCGM obtained 77.64%, and CSG obtained 80.55% precision. The proposed DCAGM obtained 93.79% precision.

5.3. Computation of Recall. The recall of a model in dance coherent action generation using deep learning is calculated by taking the ratio of the total number of correctly predicted classes to the total number of authentic classes in the dataset. It measures how many of the actual classes were predicted correctly by the model. A higher recall value indicates that the model is more likely to classify unseen data correctly. The comparison of recall has shown in the following table 5.3.

Fig. 5.3 shows the comparison of recall. In a computation tip, existing RNN obtained 62.26%, DBDL obtained 57.63%, GCGM obtained 79.42%, and CSG obtained 75.33% recall. The proposed DCAGM obtained 97.75% recall.

5.4. Computation of F1-Score. The F1-score for a Dance Coherent Action Generation Model using deep learning measures how accurately the model predicts the dancer's movement or how well the model has been trained. It is composed precision and recall. Precision measures the accuracy of the predictions made

Table 5.3: Comparison of recall (in %)

No.of Inputs	RNN	DBDL	GCGM	CSG	DCAGM
100	61.01	57.53	77.55	75.64	97.53
200	61.51	57.53	78.64	75.90	97.64
300	62.26	58.36	79.78	76.47	97.70
400	62.26	57.63	79.42	75.33	97.75
500	61.21	56.52	77.89	74.31	97.79
600	60.93	56.12	77.25	74.07	97.82
700	61.65	56.69	77.83	74.72	97.84

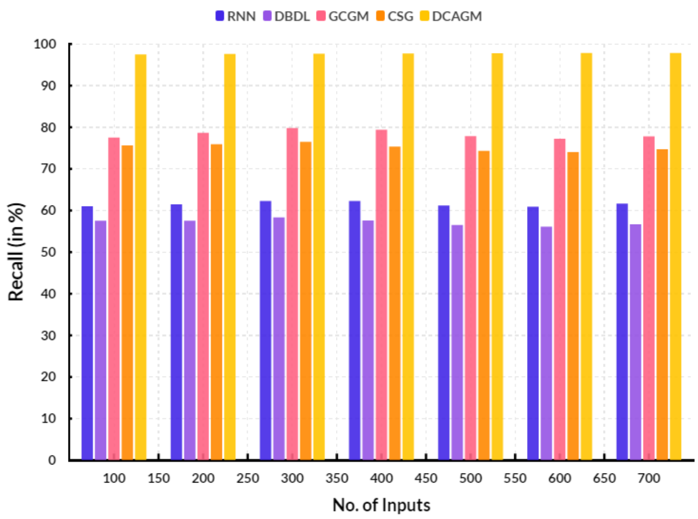

Fig. 5.3: Recall

by the model. The recall measures how well the model can identify all of the movements that the dancer is performing. By combining these two metrics, the F1 score indicates how accurately the model predicts the dancer's actual movement. The comparison of F1 score has shown in the following table 5.4

Fig. 5.4 shows the comparison of F1 score. In a computation tip, existing RNN obtained 72.46%, DBDL obtained 68.26%, GCGM obtained 74.46%, and CSG obtained 83.36% F1 score. The proposed DCAGM obtained 92.92% F1 score. The proposed dance coherent action generation model using deep learning has several advantages over traditional methods of choreography and animation:

- The model uses a deep learning approach to learn the concept of dancing from existing examples. It enables the model to capture the subtle nuances of a particular dance style, such as musicality, emotion, and the human-like movements that come with the style.
- The model can generate high-quality motions that can be repeated later with little or no modifications. It means that the motions generated by the model are less likely to drift from their representations over time, making them more suitable for use in live performances.
- The model also has the potential to adapt to changes in the environment or the music to which it is dancing, enabling the user to make minor adjustments to the dance motion generated without needing to start from scratch.

The proposed dance coherent action generation model using deep learning is limited in two main ways:

- The current deep learning models are primarily supervised and require large datasets of labeled exam-

Table 5.4: Comparison of F1 score (in %)

No.of Inputs	RNN	DBDL	GCGM	CSG	DCAGM
100	66.85	64.83	67.96	79.01	88.45
200	68.52	65.96	70.89	80.27	90.92
300	70.47	66.31	72.43	82.16	91.72
400	72.46	68.26	74.46	83.36	92.92
500	75.04	69.03	75.36	84.92	93.56
600	77.03	69.41	77.33	86.67	94.82
700	79.05	70.54	78.80	87.60	95.82

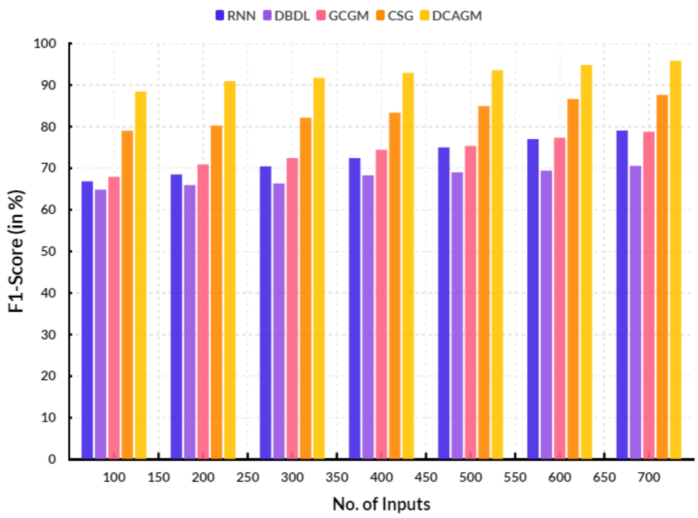

Fig. 5.4: F1 score

ples, which may be expensive and difficult to obtain.
- The models are often over-sensitive to noisy input data, making them unable to generalize and prone to over fitting.

The coherent method of the Sensor-Based Dance Coherent Action Generation Model using the Deep Learning Framework improves the existing models in several ways. Firstly, the coherent method makes use of sensors with different modalities (audio, visual, etc.) to capture the 3D coordinates (x,y,z) of each body part and to calculate the relative distances between body parts of the dancer. The data thus captured is then used to build a 3D action space model that captures motion data more efficiently and compactly compared to 2D models. This 3D action space model also improves the accuracy of the generated action sequences as it is able to capture more complex motions with greater precision. In addition, the use of deep learning algorithms to generate the action sequences helps to provide more accurate predictions while having reduced training time. Finally, the use of a real-time evaluation system helps to ensure that the generated action sequences are correct and that no mistakes are made during the generation process. All of these benefits together help to improve the existing sensor dance coherent action generation model using deep learning framework, making it more efficient, accurate, and reliable.

The proposed framework can be extended to multiple dance styles by using additional motion capture data from different dance style motions. This extended dataset could be used to train the network better and create a more generalizable model for different styles of dance. It could provide more accurate results when transferring motions from one style of dance to another. The results from this proposed framework have shown that it is possible to transfer motions between two different style movements and maintain the stylistic characteristics of the original motion. It has been achieved by preserving the continuity of the motions while modifying the amplitude and timings to fit the new style. Future research could explore various ways to improve the accuracy of this framework, including using larger datasets or adding noise to the motions to make the framework more robust. Further research could investigate the possibilities of using visuals (e.g., avatars) in addition to motion capture data to transfer both the motion and visual style of different dance styles. It could enable the framework to provide more realistic motion transfers between different dance styles.

6. Conclusions. The dance coherent action generation using deep learning is a feasible and promising approach to generating natural, realistic, and choreographically composed actions. Deep learning has enabled the autonomous generation of dance motions that are inherently choreographed and expressive. It has generated high-quality and expressive actions for various dance styles, including modern, tap, street dance, ballroom, and Latin. In addition, the generated motions demonstrate great diversity and variation, allowing for the implementation of a procedurally generated scenario in which motion is generated from a given scenario. The proposed model obtained 93.79% accuracy, 93.79% precision, 97.75% recall and 92.92% F1 score. The future scope of dance coherent action generation using deep learning is comprehensive due to the virtually limitless possibilities of combining Artificial Intelligence (AI) techniques such as deep learning with computer vision and natural language processing to create powerful machines capable of recognizing and generating complex choreography. It could be applied in various contexts, such as virtual tours, robotic dance groups, virtual reality performances, or even healthcare contexts for physical therapy and rehabilitation. With advancements in technology and imagination, the possibilities are endless.Sensor-based dance coherent action generation models using deep learning frameworks can be used in applications such as automatic motion generation and analysis. It can be applied to various robotic applications such as character animation for animated movies and games or even for industrial robotics for precise and efficient operations. For future research, this model can be used to create more complex, sophisticated, and realistic motion sequences. It can be used to develop motion generation methods that can learn from complicated interactions between different actors in the environment and for detecting and recognizing motions in a natural human motion. It can also be used to extend and improve the performance of existing motion recognition and generation methods. This model can be used to explore using motion to further improve current deep learning frameworks, especially for tasks like image segmentation and classification.

REFERENCES

[1] Ma, X., & Wang, K. "Dance Action Generation Model Based on Recurrent Neural Network," Mathematical Problems in Engineering, 2022.

[2] Kong, Y., & Fu, Y. "Human action recognition and prediction: A survey," International Journal of Computer Vision, vol. 130, pp. 1366-1401, 2022.

[3] Ushiyama, J., Takahashi, Y., & Ushiba, J. "Muscle dependency of corticomuscular coherence in upper and lower limb muscles and training-related alterations in ballet dancers and weightlifters," Journal of applied physiology, vol. 109, pp. 1086-1095, 2010.

[4] Lu, R. "Analysis of Main Movement Characteristics of Hip Hop Dance Based on Deep Learning of Dance Movements," Computational Intelligence and Neuroscience, 2022.

[5] Maithili, K., Vinothkumar, V., & Latha, P. (2018). Analyzing the security mechanisms to prevent unauthorized access in cloud and network security. Journal of Computational and Theoretical Nanoscience, 15(6-7), 2059-2063.

[6] Mei, P., Ding, G., Jin, Q., Zhang, F., & Jiao, Y. "Quantum-based creative generation method for a dancing robot," Frontiers in Neurorobotics, vol. 14, pp. 559366, 2020.

[7] Yan, S., Li, Z., Xiong, Y., Yan, H., & Lin, D. "Convolutional sequence generation for skeleton-based action synthesis," In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4394-4402, 2019.

[8] Ye, Z., Wu, H., Jia, J., Bu, Y., Chen, W., Meng, F., & Wang, Y. "Choreonet: Towards music to dance synthesis with choreographic action unit," In Proceedings of the 28th ACM International Conference on Multimedia, pp. 744-752, Oct 2020.

[9] Hu, H., Liu, C., Chen, Y., Jiang, A., Lei, Z., & Wang, M. "Multi-Scale Cascaded Generator for Music-driven Dance Synthesis," In 2022 International Joint Conference on Neural Networks (IJCNN). pp. 1-7, July 2022.

[10] Zhang, S., Yao, L., Sun, A., & Tay, Y. "Deep learning based recommender system: A survey and new perspectives," ACM computing surveys (CSUR), vol. 52, pp. 1-38, 2019.

[11] Muthukumaran, V., Kumar, V. V., Joseph, R. B., Munirathanam, M., & Jeyakumar, B. (2021). Improving network security

based on trust-aware routing protocols using long short-term memory-queuing segment-routing algorithms. International Journal of Information Technology Project Management (IJITPM), 12(4), 47-60.

[12] Zhang, X., Yang, S., Xu, Y., Zhang, W., & Gao, L. ”Mining and applying composition knowledge of dance moves for style-concentrated dance generation,” In Proceedings of the AAAI Conference on Artificial Intelligence (vol. 37, 5411-5419, June 2023.

[13] Aberman, K., Shi, M., Liao, J., Lischinski, D., Chen, B., & Cohen-Or, D. ”Deep video-based performance cloning,” In Computer Graphics Forum, vol. 38, pp. 219-233, May 2019.

[14] Zheng, D., & Yuan, Y. ”Time Series Data Prediction and Feature Analysis of Sports Dance Movements Based on Machine Learning,” Computational Intelligence and Neuroscience, 2022.

[15] Dong, Y., Su, H., Zhu, J., & Zhang, B. ”Improving interpretability of deep neural networks with semantic information,” In Proceedings of the IEEE conference on computer vision and pattern recognition , pp. 4306-4314, 2017.

[16] Sünderhauf, N., Brock, O., Scheirer, W., Hadsell, R., Fox, D., Leitner, J., ... & Corke, P. ”The limits and potentials of deep learning for robotics,” The International journal of robotics research, vol. 37(4-5), pp. 405-420, 2018.

[17] Wu, B., Xie, Z., Liang, X., Xiao, Y., Dong, H., & Lin, L. ”Image Comes Dancing With Collaborative Parsing-Flow Video Synthesis,” IEEE Transactions on Image Processing, vol.30, pp. 9259-9269, 2021.

[18] Kico, I., Grammalidis, N., Christidis, Y., & Liarokapis, F. ”Digitization and visualization of folk dances in cultural heritage: a review,” Inventions, vol. 3, pp. 72, 2018.

[19] Yin, W., Yin, H., Baraka, K., Kragic, D., & Björkman, M. ”Dance style transfer with cross-modal transformer,” In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 5058-5067, 2023.

[20] Ahmad, M., Beddu, S., binti Itam, Z., & Alanimi, F. B. I. ”State-of-the-art compendium of macro and micro energies. Advances in Science and Technology Research Journal. vol. 13, pp 88–109 https://doi.org/10.12913/22998624/103425, 2019.

[21] Cai, X., Xi, M., Jia, S., Xu, X., Wu, Y., & Sun, H. ”An Automatic Music-Driven Folk Dance Movements Generation Method Based on Sequence-To-Sequence Network,” International Journal of Pattern Recognition and Artificial Intelligence, vol. 37, pp. 2358003, 2023.

[22] Li, N., & Boers, S. ”Human Motion Recognition in Dance Video Images Based on Attitude Estimation,” Wireless Communications and Mobile Computing, 2023.

Edited by: Polinpapilinho Katina
Special issue on: Scalable Dew Computing for Future Generation IoT systems
Received: Sep 29, 2023
Accepted: Dec 13, 2023