J. Electrical Systems 20-7s (2024): 2892-2907

Xiaoyun Cao[1*],

Yibo Wang[2],

Chenghua Yin[3]

Application of Dance Rhythm Analysis and Music Matching Algorithm in the Choreography Process

Journal of Electrical Systems

Abstract: - Improvised dancing choreographies represent a crucial area of research within cross-modal analysis. Central to this endeavour is the challenge of effectively correlating using a statistical one-to-many mapping, music and dance. This mapping is instrumental in generating authentic dances across diverse genres. In this manuscript, application of dance rhythm analysis and music matching algorithm in the choreography process (DRA-MMAC-SPGAN) is projected. The pictures are gathered from AIST++ Dance Motion information set are given as input. The input images are fed to pre-processing using Sub Aperture Keystone Transform Matched Filtering (SAKTMF) for remove the background sound from the input pictures. Afterward a pre-processed picture is provided to Holistic Dynamic Frequency Transformer (HDFT) for extracting the music characteristics like onset strength envelope, Mel-frequency cepstral constants, Chroma energy normalized and peak of onset strength envelope. Then the extracted features are given to MCoCo for segmenting the music beats. In general, Semantic-Preserved Generative Adversarial Network (SPGAN) does not discuss modifying optimization techniques to identify ideal parameters to guarantee accuracy dance generated based on music. Hence, the BFO is to optimize to SPGAN which accurately generate the dance choreography based on music. The proposed DRA-MMAC-SPGAN approach is applied in Python. The presentation of the suggested DRA-MMAC-SPGAN approach attains 22.54%, 26.36% and 25.95% higher accuracy, 20.63%, 23.86% and 25.96% higher recall and0.5%, 0.7% and 0.3% lower hit rate compared with existing methods like music-to-dance motion choreography with adversarial learning (MDMC-GAN), a deep music recommendation method based on human motion analysis (DMR-HMA-LSTM-AE) and generating dances with music beats using conditional generative adversarial networks (GDMB-CGAN) respectively.

Keywords: Dance, Music, Choreography, Rhythm Of Motion, Music Beat, Background Noise, Semantic-Preserved Generative Adversarial Network, and Sub Aperture Keystone Transform Matched Filtering.

I. INTRODUCTION

In the choreography process, the application of dance rhythm analysis holds significant promise for enhancing the creative synthesis of music and movement [1-3]. Dance, as an art form deeply intertwined with music, relies on rhythmic synchronization to convey emotions and symbolism effectively [4]. Music-driven dance motion choreography seeks to harness this correlation by creating movement sequences that resonate with the periodicity and mood of the music [5]. A well-designed choreography system should exhibit creativity, generating original and meaningful dance movements in response to diverse musical inputs [6-8]. Achieving this requires a nuanced understanding of both low-level perceptual mechanisms and high-level cognitive processes involved in creativity [9]. By simulating human judgment of music-dance rhythmic correspondence and aesthetic appeal, through techniques such as deep learning models acting as discriminators, choreography systems can iteratively refine their output, incorporating feedback to enhance the created dance sequences' quality and authenticity [10]. Thus, the integration of dance rhythm analysis into the choreographic process not only enhances artistic expression but also facilitates the exploration of new creative possibilities in multimedia applications like music imagining and automatic dance formation [11-13].

The application of dance rhythm analysis in the choreography process offers several benefits, such as enhancing synchronization between music and movement and fostering creativity through informed decision-making [14, 15]. However, drawbacks exist, primarily stemming from the difficulty of precisely assembling and deciphering the subtleties of human movement [16]. Traditional methods may overlook subtleties in motion, leading to a potential mismatch between choreographed sequences and the intended rhythm of the music [17, 18]. Additionally, reliance on technology for analysis may inadvertently restrict artistic expression, as choreographers may feel constrained by rigid analytical frameworks. Moreover, there is a risk of

[1] [1*]Guangdong University of Education, Guangzhou, Guangdong, 510303, China

[1*]Email: xiaoyuncaophd@gmail.com

[2]Guangdong South China Normal University, Guangzhou, Guangdong, 510631, China

[3]Guangdong University of Education, Guangzhou, Guangdong, 510303, China

oversimplification or misinterpretation of rhythmic patterns, especially in dynamic and improvisational dance styles, which could result in choreographic stagnation or loss of authenticity [19]. Therefore, while dance rhythm analysis holds promise for enhancing choreographic processes, careful consideration of its limitations and integration with artistic intuition is essential for its effective application [20].

The motivation for applying deep learning techniques to dance rhythm analysis in the choreography process stems from the desire to enhance the creative synthesis of music and movement in street dance videos. Traditional methods for rhythm analysis may not fully capture the nuanced periodic motion changes of dancers or effectively integrate rhythm information from both visual and auditory modalities. SPGAN offers the potential to automatically extract complex patterns and features from AIST++ Dance Motion Dataset, allowing for more accurate and comprehensive analysis of motion and music. By leveraging deep learning examples, like convolutional neural networks (CNNs) for visual characteristic extraction and recurrent neural networks (RNNs) for chronological sequence analysis, can develop systems capable of detecting and aligning rhythm patterns between dance movements and music beats. This approach enables the generation of original and meaningful dance sequences that are synchronized with high-quality music, facilitating the creation of immersive music videos and enhancing the overall aesthetic appeal of multimedia presentations. Additionally, Semantic-Preserved Generative Adversarial Network (SPGAN) techniques can adapt and learn from feedback, enabling iterative refinement of choreographic output based on human judgments of rhythmic correspondence and aesthetic quality. Overall, incorporating SPGAN- BFO into dance rhythm analysis holds promise for pushing the boundaries of creativity and expression in choreography, ultimately enriching the artistic experience for both creators and audiences.

The main contributions of this research work are summarized below

- In this research, application of dance rhythm analysis and music matching algorithm in the choreography process (DRA-MMAC-SPGAN) is proposed.
- Develop a Sub Aperture Keystone Transform Matched Filtering (SAKTMF) founded pre-processing technique for eliminate the background sound from the images. Unpaired Multi-View Graph Clustering (UMGC) has developed to segment the Region of Interest (ROI) of the input image.
- Music Features are extracted using Holistic Dynamic Frequency Transformer (HDFT). Semantic-Preserved Generative Adversarial Network (SPGAN) is constructed for generate the dance choreography based on music.Propose a BFO to enhance the weight limit of SPGAN.
- The effectiveness of the suggested model is examined using current techniques like MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN models respectively.

The remaining manuscripts are arranged as follows: Part 2 provides a brief overview of the previous study, Part 3 outlines the suggested technique, Part 4 discusses and confirms the findings, and Part 5 wraps up the paper.

II. RECENT RESEARCH WORK: A BRIEF REVIEW

A number of previously published studies in literature relied on deep learning-based dance rhythm analysis for choreography. Few of them were mentioned here.

Guofei Sun et al. [21] introduced Deep Dance, a GAN-based cross-modal association framework that associated dance motion and music, two distinct modalities, in order to produce the desired dance sequence in terms of the input music. Using examples as a guide, its generator forecast which dance move would go best with the present musical composition. However, its discriminator assessed the entire performance in the capacity of an outside audience member. If the discriminator was unable to discern the generated motions from the training samples, then the generated dance movements and the matching input music were deemed well-matched, based on the estimated likelihood. Long, realistic dancing sequences were produced by the provided system by including motion consistency restrictions into their loss function. The challenge of costly and ineffective data gathering was addressed by their useful method of turning an open data source into YouTube-Dance3D, a sizable dataset. It provides high accuracy and low recall.

Wenjuan Gong and Qingshuang Yu [22] created a deep learning music recommendation system using dance motion analysis, and assessed it using quantitative metrics. This work built an LSTM-AE based music recommendation algorithm that learns the correspondences between motion and music for quantitative evaluation. It provides high precision and low f-measure.

Yin-Fu Huang and Wei-De Liu [23] introduced an adversarial network-based conditional generative choreography system using music input. Using MFCC characteristics and dance Skeletons taken from Japanese dance videos, they first created the MFDS dataset. The dancing skeletons were identified by analyzing the picture frames of a film, and the MFCC characteristics were retrieved based on musical beats. The music-driven choreography system was trained using a generative adversarial network during the training phase. The discriminator used traditional CNNs, while the generator combined residual blocks into fractionally stridden convolution. It provides high hit rate and low accuracy.

Kosmas Kritsis et al. [24] introduced a multimodal convolutional autoencoder that generated unique dancing motion sequences of any duration by fusing together 2D skeletal and audio data using an attention-based feature fusion technique. Using solely skeletal data as input, they first verified that the system could capture the temporal context of dance in anunmoral scenario. In the first user research, 24 users submitted 1440 rating replies, which demonstrated that the model trained with 500 posture input sequences had the best performance. Based on these results, they taught the multimodal architecture that was given using two distinct methods to address the autoregressive mistake accumulation phenomenon: self-supervised curriculum learning and teacher-forcing. Low recall and a high genre score are provided.

Doyoung Kim et al. [25] developed the Dance Quality Assessment (DanceQA) Framework to assess dance performance, taking into account choreographic elements that were crucial DanceQA subjective criteria. From the standpoint of human perception, they discovered that kinematic variation and rhythmic alignment were important choreographic elements. Two measures, kinematic information entropy (KIE) and kinematic-music beat similarity (BSIM), were developed based on these variables. This study showed that there was a strong correlation between these measurements and particular body parts in each dance. It provides high precision and low genre score.

Changrui Cui et al. [26] presented a technique for using motion capture data to automatically generate Labanotation. First, they transformed the BVH-formatted Euler angle data into a series of 3D Cartesian world coordinates based on the examination of human bones in order to make the computation of the bending angle and motion feature sequences easier. Second, they demonstrated a technique for segmenting motion based on kinematics traits and rhythms that might break down various motion types into smaller pieces for additional identification. Low genre score and high f-measure are its features.

Shanshan Kong [27] developed a deep learning toddler dance generation model that extracts dance and music characteristics depending on beat and rhythm of the song. Furthermore, the research produced fluid dances via a generator module, strengthened the correspondence between the dances and music produced by the model via a discriminator, and improved the audio features' representativeness using a self-encoder module. It provides freshet inception distance low hit rate.

III. PROPOSED METHODOLOGY

In this sector, application of dance rhythm analysis and music matching algorithm in the choreography process (DRA-MMAC-SPGAN) is deliberated. Block diagram of suggested DRA-MMAC-SPGAN technique is in Figure 1. It covers such stages as Sub Aperture Keystone Transform Matched Filtering, Holistic Dynamic Frequency Transformer, Multi-Level Consistency Collaborative Multi-View Clustering, Semantic-Preserved Generative Adversarial Network and Bitterling Fish Optimization.

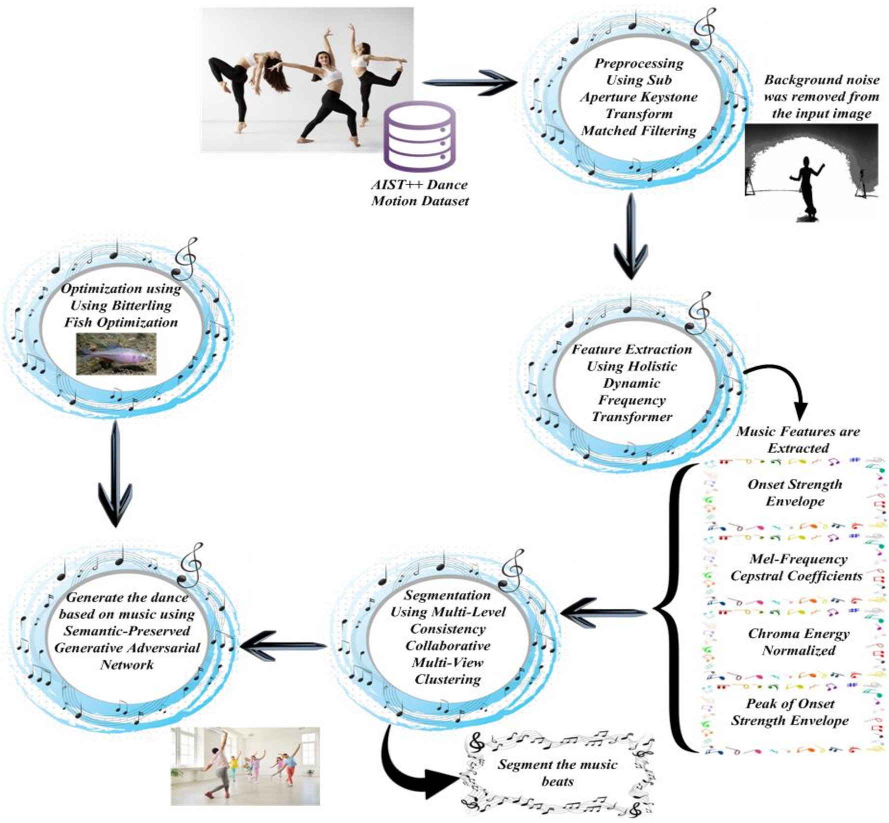

Figure 1: Block diagram of proposed DRA-MMAC-SPGAN

A dataset AIST++ Dance Motion Dataset integrating during the data pre-processing stage, dance videos gathered from the AIST Dance Video DB are used to build music features. Beats from a video's visual frames are used to extract the music's characteristics. Lastly, arrange them in the AIST++ Dance Motion Dataset in an aligned fashion. During the training stage, train the choreographic system using a Semantic-Preserved Generative Adversarial Network. The generator and discriminator are two separate components of the GAN. Based on prior motion, the noise from a normal distribution and the elements of the music that are now playing, the generator creates motion. The discriminator modifies the discriminator's parameter weights for the generated phase by determining if it is continuous with the preceding motion and fits the current music. In addition, the discriminator uses the genuine skeleton as input to determine the loss. During the online stage, the trained generator may provide the appropriate dance movements based on the music. Consequently, a thorough description of each step is provided below.

A. Image Acquisition

The input image are gathered from AIST++ Dance Motion Dataset [28]is built using the AIST Dance Video Database. An intricate pipeline is created to estimate the camera parameters, 3D human keypoints, and 3D human dance motion sequences from multi-view recordings. It offers camera settings and 3D human keypoint annotations for 10.1M photos, encompassing 30 distinct subjects in 9 viewpoints. It is the biggest and richest dataset with 3D human key point annotations currently available because of these characteristics. In addition, 1,408 sequences of 3D human dancing motion are included, which are shown as joint rotations and root trajectories. Ten dance genres with hundreds of choreographies each have an equal distribution of dancing actions. Motion times range from 7.4 to 48.0 seconds. Every dancing move has an associated piece of music.

B. Pre-Processing Using Sub Aperture Keystone Transform Matched Filtering

In this section, Sub Aperture Keystone Transform Matched Filtering [29] (SAKTMF) is used for preprocessing the data. SAKTMF is used to remove the background noise from the image. The advantages of SAKTMF make it an ideal choice for dance rhythm analysis in the choreography process, offering precision, customization, noise reduction, time efficiency, creativity enhancement, and seamless integration with existing technology. The coherent integration gain between various sub apertures enhances the SAKTMF method in virtue. SAKTMF is a sophisticated computational technique that finds application in the realm of dance rhythm analysis within the choreography process. A unique SAKTMF technique is developed depend on the phase connection derivation among sub apertures. Thus the aperture is given in equation (1)

$$PJ(h_t, h_n) = \sum_{l=1}^{L} PJ_l(h_t, h_{n,l}) \tag{1}$$

Where, PJ signifies the filtered horizontal movement at frame, (h_t, h_n) denotes a demodulation operation, L represents the sub blocks and l denotes the sub aperture index. In the generation phase, the same process is used to generate labels for dance background music input by the user. In these methods to demonstrate that distinct parts of the body often harmonize music at different musical stages is shown in equation (2)

$$pJ_l(h_t, f_n) = CA_{hn}\left[PJ_l(h_t, h_n, l)\right] \tag{2}$$

Here, $CA_{hn}(.)$ represents the performing operation through f_n, h_n signifies doppler frequency. The process of integrating sub apertures incoherently carried out and it is given by the equation (3)

$$K\{\alpha\} = \sum_{l=1}^{L} \left| pJ_l\left(\frac{2}{a} t(\alpha, lF_a), f(\alpha, lF_a\right) \right| \tag{3}$$

Here, α represents searching parametric space vector, $t(\alpha, lF_a), and\ f(\alpha, F_a)$ represents range along Doppler position corresponding to sub aperture index l. Based upon series expansion around the range frequency variable is given by the equation (4).

$$\left(\frac{d_a}{d_t + d_a}\right)^k \approx 1 - k\frac{d_t}{d_a} \tag{4}$$

Where, d_t represents the range frequency variable, h_a denotes the preliminary integration, k indicates the linear range. The data in the test group is also used randomly paired music and motion seeds. Because users may use randomly paired music and dance motions as input in practice, the test dataset also contains the randomly paired music and dance motions. The motion parameters closely correspond to the real value, and the l sub aperture signal after multiplying it is possible to express the phase compensation function using

$$Pj_{l,comp}\left(f_t, \tau_{n,l}\right) = PJ\left(f_t, \tau_{n,l}\right)\exp\left[-j\frac{4\pi}{\lambda} z_1 l F_a\right]$$

Where z denotes the searching motion parameters. It guarantees the completeness and accuracy of the data. Thus the SAKTMF of the image is given by the equation (5)

$$\hat{y}_1 = -\frac{2}{\lambda}\left[Arg\max\{SAKTMF(k_t, h_n; \hat{z})\}\right] \tag{5}$$

Here, $\hat{z}$ represents the corresponding searching motion parameters vector, k_t denotes the original attribute and λ represents the noise variable. Finally the input pictures are pre-processed successfully by eliminating the sound from the input pictures using SAKTMF. The pre-processed image is fed into Holistic Dynamic Frequency Transformer for the music extraction process.

C. Music Feature Extraction Using Holistic Dynamic Frequency Transformer (HDFT)

In this step, music feature extraction using Holistic Dynamic Frequency Transformer (HDFT) is discussed [30]. HDFT is utilized to extract the music characteristics like on set strength envelope, Mel-frequency cepstral constants, Chroma energy normalized and peak of onset strength envelope from the pre-processed output. This approach successfully captures long-range dependencies and reduces computational complexity, addressing the drawbacks of local window based attention while maintaining local structures and using less memory. Enhancing the grading process may be achieved by extracting characteristics from the photos. The Sash The attention mechanism, which is a variation where calculations are limited inside local windows that are segments of the input picture, is represented by equation (6).

$$A = soft\max\left(QK^T / \sqrt{d_K}\right)V \tag{6}$$

Where Q, K and V means the inquiry, key, and value matrices respectively. A Represent the attention map output and d_K represent the key vectors and query's dimensionality. Attention's computational complexity is a constraint. The amount of calculations needed rises quadratic ally with image size. Using window attention, complexity is decreased..The onset strength envelope feature identifies the points in time where there is a significant change in the audio signal's energy, indicating the beginnings of musical events or notes. It's useful for tasks like beat tracking and rhythm analysis is given in equation (7).

$$M_f = Q_f \,\Theta\, K_f \tag{7}$$

Where, Θ denotes the element wise multiplication. After attention, the value and attention matrix are subjected to a hadmard product to get the final output. Mel-frequency in a manner, the cepstral coefficients characteristic indicates a sound's short-term power spectrumthat's more closely related to human auditory perception. They are often used for tasks like speech and music recognition is given in equation (8).

$$A = soft\max(M)\,\Theta\, V \tag{8}$$

Finally the attention map is convolved using convolution and applied to the original input. Chroma energy normalized feature identifies the distribution of energy across the various classes in an octave. This feature normalizes the chroma vectors to reduce the impact of varying loudness levels and make the representation more robust is given in equation (9).

$$X' = CONV_{1\times1}(A) + X \tag{9}$$

Where X denotes the original input in image. Peak of onset strength envelope feature identifies the highest points in the onset strength envelope, indicating the most prominent onsets or musical events. It can be useful for tasks like segmenting music into different sections or identifying key moments in a piece. Finally, Holistic Dynamic Frequency Transformer (HDFT) has extracted the music characteristics like onset strength envelope, Mel-frequency cepstral constants, Chroma energy normalized and peak of onset strength envelope. After completing following feature extraction, the features are loaded into collaborative multi-view clustering with multi-level consistency..

D. Segmentation Using Multi-Level Consistency Collaborative Multi-View Clustering (MCoCo)

In this section segmentation using MCoCo [31] is discussed. The image pre-processing is now used to further segmentation and this paper offers a unique approach is called MCoCo. The entire dance sequence is divided into segments. MCoCo is used to segment the musical beats. The original image may contain some duplicate information, and the size and input formats of the many perspectives of a multi-view image are typically somewhat varied. For MCoCo, the whole contrastive learning loss of semantic consistency may be represented by the following equation. (10)

$$L_{Se} = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1, j\neq i}^{m} l(i, j) + \sum_{i=1}^{m}\sum_{c=1}^{k}\left(\frac{1}{N}\sum_{j=1}^{N} S_{jc}^{(i)} \log \frac{1}{N}\sum_{j=1}^{N} S_{jc}^{(i)}\right) \tag{10}$$

Where L_{Re} represents the loss of semantic constancy and $l(i, j)$ specifies to semantic constancy loss. m Represent the number of views, N represent the amount of samples and $S_{jc}^{(i)}$ represent a sematic label in ith view. The original dance within the database is synchronized with the corresponding music, and dancers initiate or conclude actions at the time of musical beats. Consequently, musical beats serve as indicators to detect the boundaries of a dance segment which can be expressed as given in the equation (11)

$$Q_{i,j}^{(m)} = \frac{\left(1 + \left\|Z_i^{(m)} - \mu_j^{(m)}\right\|^2\right)^{-1}}{\sum_j \left(1 + \left\|Z_i^{(m)} - \mu_j^{(m)}\right\|^2\right)^{-1}} \tag{11}$$

Here $Q_{i,j}^{(m)}$ represents the pseudo label, $\mu_j^{(m)}$ represent the cluster centroids and $Z_i^{(m)}$ represent the dimensional feature space. To improve the pseudo label's capacity for discriminating with more assurance, improve $Q_{i,j}^{(m)}$ to an auxiliary target transfer $P^{(m)}$ with the operation of square and standardization for MCoCo can be stated as the given equation (12)

$$P_{ij}^{(m)} = \frac{\left(Q_{ij}^{(m)}\right)^2 / \sum_i Q_{ij}^{(m)}}{\sum_j \left(\left(Q_{ij}^{(m)}\right)^2 / \sum_i Q_{ij}^{(m)}\right)} \tag{12}$$

Where L_{Se} represents the loss of semantic consistency and L_{Ml} denoted as multi-level consistency loss. We propose a new multi-level consistency collaboration strategy that can achieve multi-level collaboration by having multi-views work together to get consistent cluster assignments while also using the aligned semantic labels to weakly supervise the cluster assignments in feature space. The multi-level consistency loss L_{Ml} for MCoCo can be expressed as the given equation (13)

$$L_{Ml} = \sum_{k=1}^{m}\sum_{c=1}^{m}\left(D_{kl}\left(P^{(c)} \| Q^{(k)}\right)\right) + D_{kl}\left(S'^{(k)} \| Q^{(k)}\right) \tag{13}$$

Where L_{Ml} represent the multilevel consistency loss, $S'^{(k)}$ represent the semantic space target distribution and D_{kl} represent the divergence. The dance segments offer an advantage over poses as units for choreography, as kinematic beats consistently synchronize with the musical beats within the same dance segment. To ensure that precise cluster assignments are realized with high confidence and to prevent the interference of a few incorrect predictions, the final cluster assignment for MCoCo can be expressed as the given equation (14)

$$Y_i = \arg\max_j \left(\frac{1}{m}\sum_{k=1}^{m} Q_{ij}^{(k)}\right) \tag{14}$$

Where $P_{ij}^{(m)}$ denoted as auxiliary target distribution and $Q_{ij}^{(m)}$ is viewed as a pseudo tag that expresses the likelihood of transmission the i^{th} sample from the m^{th} view to the j^{th} group with the process square and normalization, Y_i represent the cluster assignment. To boost the more confident pseudo label's capacity to discriminate, improving to an auxiliary target distribution. Finally a MCoCo has segmented the music beats. Then the segmented output is given to Semantic-Preserved Generative Adversarial Network for generate dance based on music.

E. Choreography Generation Using Semantic-Preserved Generative Adversarial Network

In this section, SPGAN [32] is discussed. SPGAN is utilized to generate choreography based on music. The network-level disparity between sources, target domains is minimized by source-to-target translation with sematic preservation, which is accomplished by the SPGAN. The objective is to map source domain photos to target domain images as ground-truth labels are only accessible in source domain. A ground-truth dance motion so that they can generate different dances based on the same music and different motion seeds. The loss function is given in equation (15)

$$X_S = K_{zr} \sim Z_R[\log H_R(z_r)] \tag{15}$$

Where, K_{zr} denotes set of crushed-truth pixels, Z_R signifies set of foretold pixels, X_S is the loss purpose H_R is the segmentation model and z_r is the source boundary. To adjust the length of the dance segments to accommodate different musical rhythms so that the same clip of background music can be matched with more dance moves. The architecture is multi-layered, with effective categorization between the levels. Then, it is given in equation (16)

$$X_{ij} = B_{zd} \sim Z_d\left[\left\| P_\upsilon^t(C_{ij}(z_d)) - z_d \right\|_1\right] \tag{16}$$

Where, X_{ij} is the source domain, B_{zd} is the target domain, Z_d is the parameter of network Generator, $P_{\upsilon}^{\prime}$ is the rise bias, C_{ij} is the abstract constant. After the input is processed by the hidden layer, it is sent to fully connected neurons, which produce output. SPGAN is used in the proposed study for the purpose of categorization. Moreover, SPGAN resolves the problem of gradient vanishing with an explicit memory unit. The SPGAN is given in equation (17)

$$W_v = T_{zr} \sim Z_R \left[N_{enc}(z_r) - F_{enc}(Z_R) \right]_1 \tag{17}$$

Where, W_v is the stylized images, T_{zr} is the semantic content, Z_R is the set of predicted pixels, N_{enc} is the mixed images, z_r is the source boundary and F_{enc} is the reconstruction error. Furthermore, SPGAN allows for the use of any weights to memorize anything. The generator and discriminator are optimized in opposite ways during training. Only the training phase makes use of the discriminator. In the testing phase, the trained generator takes any piece of music and creates a dance routine with the same duration.

$$E_j = Q_{(y_d - z_d) \sim (Y_d - Z_d)} \tag{18}$$

Where, E_j denotes for probability predicted by segmentation method and Q hyper parameter, y_d is the cell stage, z_d is the input gate, Y_d is the candidate state value and Z_d is the gate activation. The music beats produced by drum form the dominant rhythm of music. They are strong and repeat with a fixed period. If the speeds of two music pieces are the same, their dominant rhythms are identical.

$$M_{total} = M_s + \lambda_v \cdot (M_{rc\_c} + M_{mix\_c}) \tag{19}$$

Where, M_s is the bias term, λ_v is the logistic function, M_{rc_c} is the Varlet discretization and M_{mix_c} is the vector parameters, M_{total} is the total bias vector. Beats in music are identified by integrating energy dynamics across several frequency ranges. Then, the term "periodically evolved beats" is employed to characterize musical rhythm. The choreography generator may produce the appropriate dance movements based on the provided music. Finally, by using SPGAN generated the dance based on music. Here, BFO is employed to tuning weight, bias parameter for optimize the SPGAN.

F. Optimization Using Bitterling Fish Optimization (BFO)

In this section, the optimization using BFO[33] is discussed. The Bitterling Optimization Algorithm (BOA) is a metaheuristic approach inspired by the reproductive behavior of bitterling fish, wherein males compete for access to suitable spawning hosts. BOA demonstrates superior accuracy in solving NP-Hard problems compared to traditional metaheuristic methods, is unique balance of exploration and exploitation strategies derived from natural selection principles. By mimicking the competitive selection process of bitterling fish, BOA reduces the risk of converging to local optima while efficiently exploring the solution space, making it a promising choice for diverse optimization tasks.

1) Stepwise Procedure for BFO

The step by step process for obtaining appropriate SPGAN values using BFO is described here. To creates a uniformly distributed population for enhancing ideal SPGANparameters. The entire step is presented in below,

Step 1: Initialization Phase

The solution to the problematic is a bitterling fish or egg. Several bitterling fish inhabitants are shaped randomly, as shown in Equation (20).

$$B = \begin{bmatrix} B_1^1 & B_1^2 & \dots & B_1^C \\ B_2^1 & B_2^2 & \dots & B_2^C \\ \dots & \dots & \dots & \dots \\ B_n^1 & B_n^2 & \dots & B_n^C \end{bmatrix} \tag{20}$$

Where C is the amount of dimensions or choice variables for each solution. The assessment of B is a matrix prime populations value of n .

Step 2: Random Generation Phase

Parameters generated at random for input. The selection of ideal fitness values was based on a clear hyperparameter condition..

Step 3: Fitness Function

It creates unplanned solution from reset values. It has calculated by optimizing parameter. Then the formula is derived in equation (21).

$$Fitness\ function = Optimizing\left\lfloor P_{\upsilon}^{t}, E_{j}\right\rfloor \tag{21}$$

Where P_{υ}^{t} is used for increasing the accuracy and E_j is used for reducing hit rate.

Step 4: Search and Seize Oysters for Optimizing P_{υ}^{t}

Any fish or solution can scour the problematic area for suitable oyster mating sites. According to the suggested plan, every fish is located in an area with better-quality shells. A fish can search for oysters and approach them. This fish captures the target oyster after another fish fails to notice it. Equation (22) describes the state of oyster possession:

$$V_i^{t+1} = \begin{cases} K.\,V_i^t + (F^{+} - K.\,V_i^t)..\zeta\ r \le P_{\upsilon}^{t} \\ K.\,V_i^t + (F^{*} - K.\,V_i^t)..\zeta\ r > P_{\upsilon}^{t} \end{cases} \tag{22}$$

Where V_i^t and V_i^{t+1} represent the current and new position of the fish, F^* represent the best oyster and F^+ represent the one of the worthies of the oysters' population. ζ and r represent the random number, and K represent the fish moves to seepage the oyster.

Step 5: Escape and Not Seize the Oyster for Optimizing E_j

The behavior known as "Escape and not seize the oyster" occurs when a different fish looks after the oyster, leading the other fish to avoid getting close to it and instead take up a different position. Equation (23) is utilized to go away from or inadvertently look for a fish that hasn't been able to catch oysters.

$$V_i^{t+1} = \begin{cases} K.\,V_i^t + (F^{*} - K.\,E_j)..\zeta\ r \le 0.5 \\ \quad l + (u - l).\ .\zeta\ r > 0.5 \end{cases} \tag{23}$$

Where, (I) is the input node features and $(f(F_j^t))$ is the impartial function value for this answer.

Step6: Termination

A weight limit value of producer P_{υ}^{t} and E_j from SPGAN is enhanced by utilizing BFO; and it will repeat step 3 until it obtains its hesitant criteria $B = B + 1$. Then DRA-MMAC-SPGAN assesses to generate the dance by increasing the accuracy and Lessing the calculation time.

IV. RESULT AND DISCUSSION

Outcomes of the future DRA-MMAC-SPGAN technique have generated the dance choreography based on music. This suggested technique is implemented using Python and assessed by using several performance analysing metrics like Accuracy, precision, recall, F-measure, dance similarity, frechet inception distance, hit ate and genre score are analysed. The results of the proposed DRA-MMAC-SPGAN practice are compared to those current methods like MDMC-GAN [21], DMR-HMA-LSTM-AE [22] and GDMB-CGAN [23]. Table 1 illustrate the output result of proposed DRA-MMAC-SPGAN.

Table 1: Output result of proposed DRA-MMAC-SPGAN

Dataset	*Input Image*	*Preprocessed Image*	*Segmented Image*
AIST++ Dance Motion Dataset			

A. Performance Metrics

Accuracy, precision, recall, F-measure, dance similarity, FID, hit score, and genre score are some examples of performance metrics. It is determined to scale the performance parameters using the confusion matrix.

1) Accuracy

The ratio of the number of samples correctly classified by scheme to the total number of samples is used to compute the accuracy value which is computed using equation (24)

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \tag{24}$$

Where TP represent the true positive, TN represent the true negative, FP represent the false positive and FN represent the false negative.

2) Precision

In other words, it assesses the predictive capability of the sample by analyzing its predictive value, which may be either positive or negative depending on the class for which it is computed, which is definite by equation (25).

$$\Pr ecision = \frac{TP}{(TP+FP)} \tag{25}$$

3) Recall

The recall of a machine learning model measures how well it can recognize good samples. Put another way, it measures the likelihood of getting a favourable result. That's provided in equation (26)

$$\mathrm{Re}call = \frac{TP}{TP+FN} \tag{26}$$

4) F-Measure

The F1 score or F-measure are other names for the F-score. It is a statistic for assessing how well a machine learning model is doing. It generates a single score by combining recall and accuracy.

$$F-Measure = \frac{2*(precision*recall)}{precision+recall} \tag{27}$$

5) Frechet Inception Distance

A measure called the FID score determines the separation between feature vectors computed for generated and actual pictures.

6) Hit Rate (HR)

After the clusters are defined, if a produced dance intersects the ground-truth cluster, HR is set to 1; if not, HR is set to 0.

7) Genre Score (GS)

If a produced dance falls under the ground-truth dancing genre, the genre score is 1; if not, GS is 0.

B. Performance Analysis

The imitation outputs of DRA-MMAC-SPGAN method are portrayed in figure 2 to 8. The future DRA-MMAC-SPGAN approach is compared to existing MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN models.

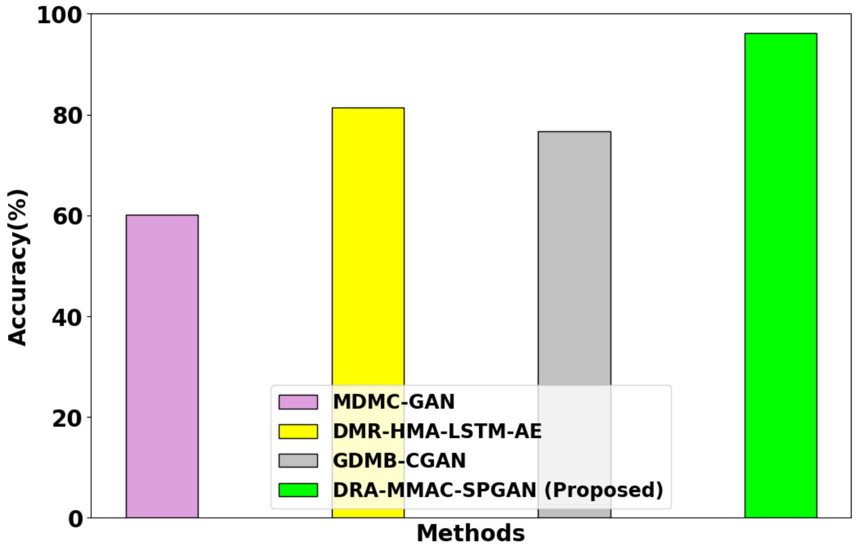

Figure 2: Performance analysis of accuracy

Figure 2 shows the presentation analysis of Correctness. Dance rhythm analysis helps choreographers break down music into its rhythmic components, such as beats, accents, and tempo variations. By understanding these elements, choreographers can synchronize movements precisely with the music, ensuring that each step aligns with the rhythm using proposed DRA-MMAC-SPGAN method. The proposed DRA-MMAC-SPGAN method attains 22.54%, 26.36% and 25.95% higher Accuracy as compared to the existing methods MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN respectively.

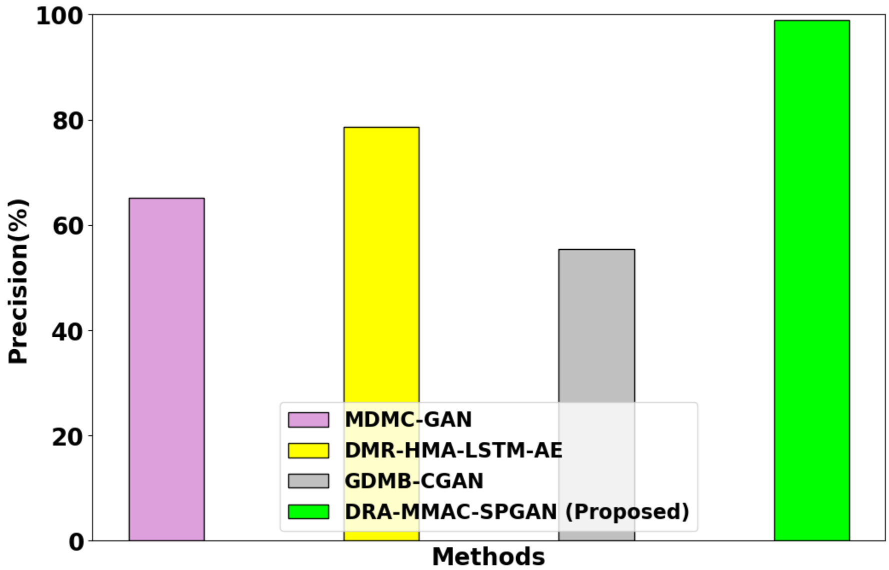

Figure 3: Performance analysis of precision

Figure 3 shows the presentation examination of precision. Dance rhythm analysis involves dissecting the musical composition to grasp its underlying structure, including beats, tempo, accents, and phrasing. Choreographers can use this understanding to synchronize movements with the music, ensuring coherence and harmony between sound and motion. The proposed DRA-MMAC-SPGAN method attains 23.43%, 26.32% and 25.92% higher precision as likened to the current methods MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN respectively.

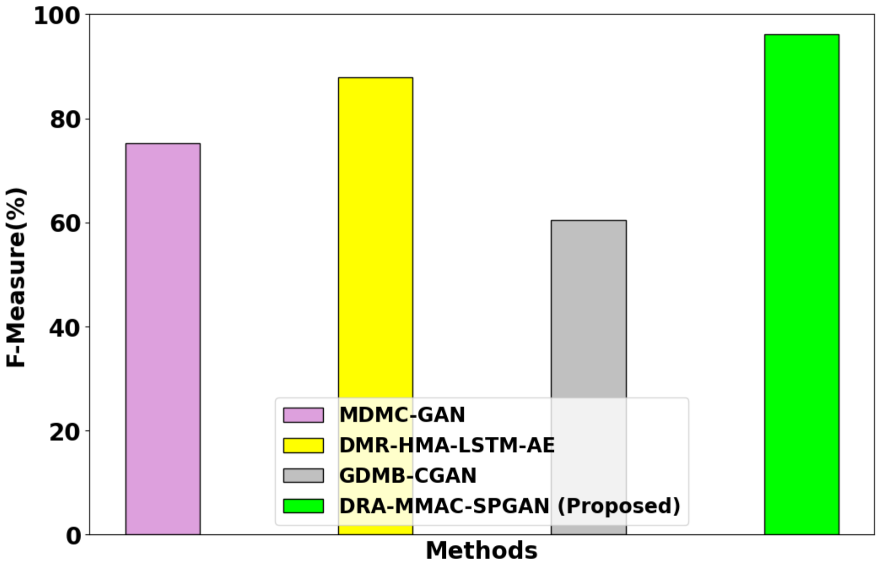

Figure 4: Performance analysis of F-measure

Figure 4 shows the presentation examination of F-measure. It's particularly useful when you have uneven class distribution, which might be the case when analysing dance rhythms where certain rhythms or patterns might be more prevalent than others. A high F-measure indicates that the choreography process is both accurate in identifying desired rhythms or patterns and comprehensive in capturing all relevant instances within the music. The proposed DRA-MMAC-SPGAN method attains 21.56%, 24.35% and 25.98% higher F-measure as compared to the current methods MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN respectively.

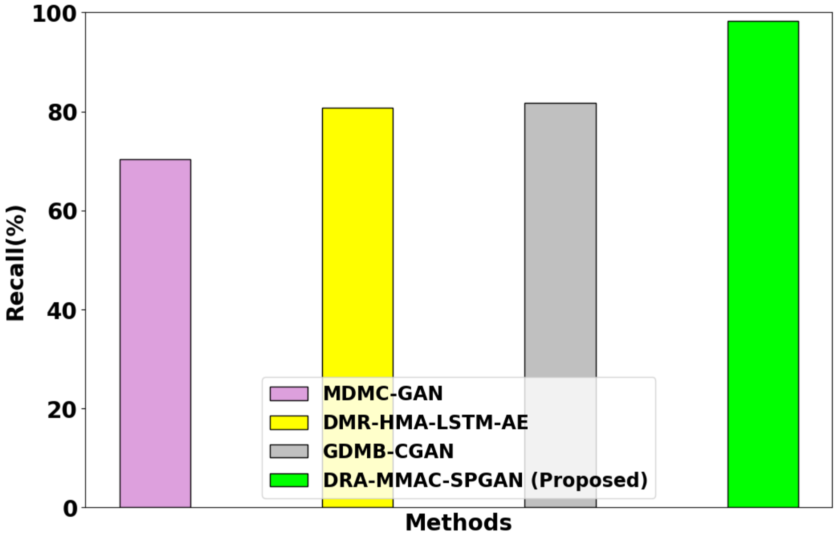

Figure 5: Performance analysis of recall

Figure 5 shows the recital analysis of recall. A memory graph could visually depict the sequence of dance movements in a choreographic piece. Each node in the graph represents a specific movement or step, and the edges between nodes indicate the transitions or connections between these movements. The proposed DRA-MMAC-SPGAN method attains 20.63%, 23.86% and 25.96% higher recall as likened to the present approaches MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN respectively.

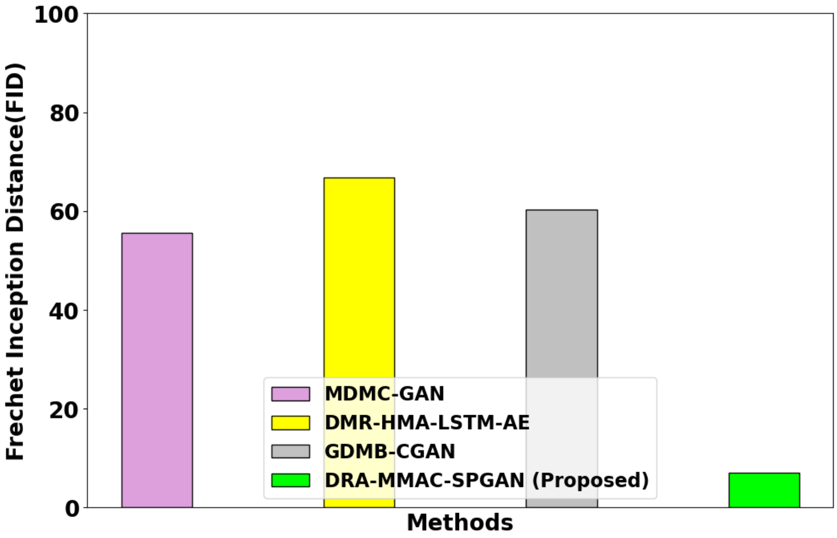

Figure 6: Performance analysis of frechet inception distance

Figure 6 shows the presentation examination of FID. Frechet Inception Distance (FID) is a metric commonly utilized in evaluating the excellence of pictures made by generative models, particularly within the domain of computer vision. A lower FID designates a higher similarity among the produced sequences and the ground truth, suggesting that the choreography process is effective in producing sequences that resemble professional choreography. Conversely, a higher FID suggests that the generated sequences diverge significantly from the expected quality, indicating areas for improvement in the choreography process. The proposed DRA-MMAC-SPGAN method attains 22.65%, 24.73% and 15.01% lower FID as compared to the existing methods MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN respectively.

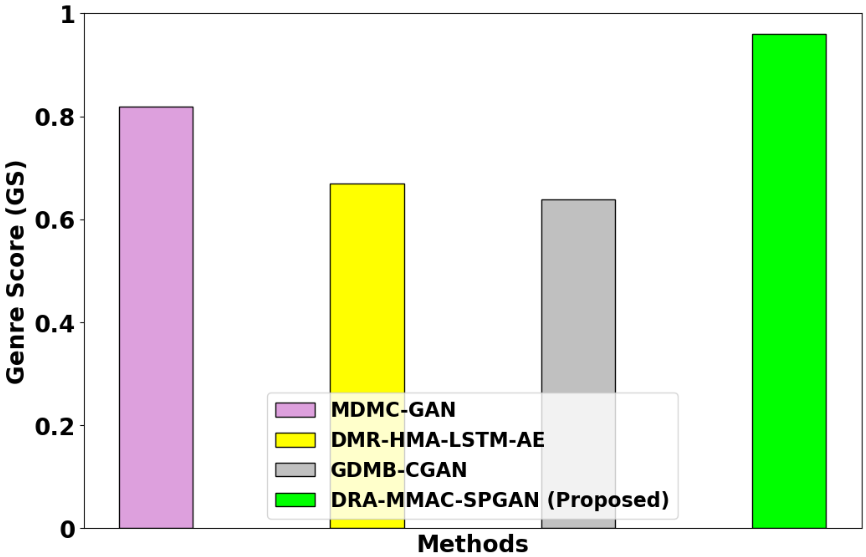

Figure 7: Performance analysis of genre score

Figure 7 shows the performance examination of genre score. Genre score analysis in the choreography process involves a comprehensive examination of the elements that define a particular dance genre, including rhythm, musicality, movement vocabulary, historical context, and emotional expression. By understanding these elements in depth, choreographers can create choreography that is both technically proficient and artistically meaningful within the context of the chosen genre. The proposed DRA-MMAC-SPGAN method attains 0.76%, 0.58% and 0.97% higher genre score as compared to the existing methods MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN correspondingly.

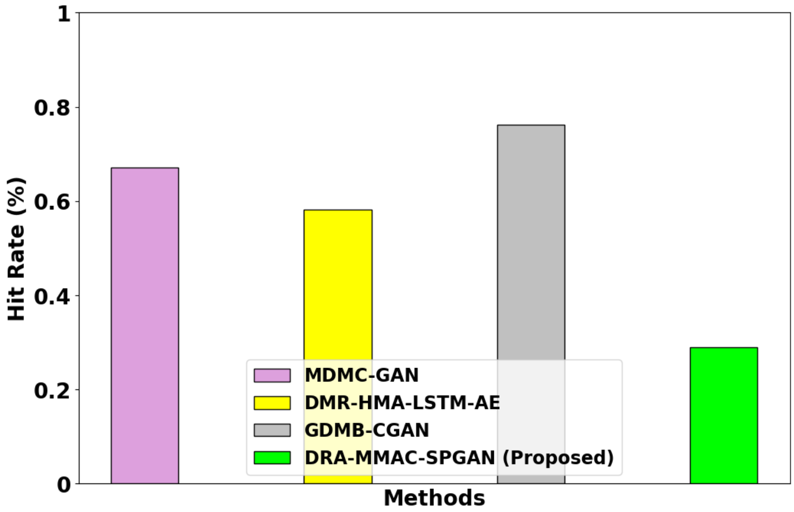

Figure 8: Performance analysis of hit rate

Figure 8 shows the performance analysis of hit rate. Hit rate refers to the accuracy with which dancers execute movements in alignment with the rhythm of the music. Hit rate analysis informs the training and rehearsal process for dancers. Choreographers and dance instructors use it to provide feedback to dancers, helping them improve their timing, coordination, and musical interpretation. The proposed DRA-MMAC-SPGAN method attains 0.5%, 0.7% and 0.3% lower hit rate as compared to the existing methods MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN respectively.

C. Discussion

The paper introduces DRA-MMAC-SPGAN, a novel application of dance rhythm analysis with a music matching algorithm for choreography. DRA-MMAC-SPGAN excels across various performance metrics, showcasing its efficacy in generating dance movements harmonized with music. Through dance rhythm analysis, it precisely dissects music into rhythmic components, ensuring synchronization of movements with the music's structure. The method outperforms existing techniques in accuracy, precision, F-measure, recall, and genre score analyses, achieving improvements of 22.54%, 23.43%, 21.56%, 20.63%, and 0.76%, respectively, over MDMC-GAN; 26.36%, 26.32%, 24.35%, 23.86%, and 0.58% over DMR-HMA-LSTM-AE; and 25.95%, 25.92%, 25.98%, 25.96%, and 0.97% over GDMB-CGAN, respectively. Moreover, DRA-MMAC-SPGAN achieves lower FID, signifying its capability to produce choreography sequences resembling professional standards, with reductions of 22.65%, 24.73%, and 15.01% compared to MDMC-GAN, DMR-HMA-LSTM-AE, and GDMB-CGAN, respectively. Despite minor variations in hit rates, DRA-MMAC-SPGAN holds promise in providing accurate feedback for dancers to enhance timing and coordination. Overall, the results underscore the transformative potential of DRA-MMAC-SPGAN in revolutionizing choreography by seamlessly aligning dance movements with musical rhythms.

V. CONCLUSION

In this manuscript, application of dance rhythm analysis and music matching algorithm in the choreography process (DRA-MMAC-SPGAN) was successfully implemented. A unique generator is shown that uses a beginning dancing stance and a music sequence as input to create dance sequences frame-by-frame. To capture the relationships between music and dance pairings and enforce the generator to produce dance sequences with distributions like the real data, a conditional discriminator is suggested. Add motion consistency restrictions to the loss function and use a differential forward kinematic operator to facilitate simple reverse propagation of loss gradients in order to better produce realistic dancing motions. SPGAN-BFO is used to produce the dance movements based on music. The proposed DRA-MMAC-SPGAN approach is applied in Python. The presentation of the proposed DRA-MMAC-SPGAN approach contains 22.54%, 26.36% and 25.95% higher accuracy, 21.56%, 24.35% and 25.98% higher F-measure and 0.5%, 0.7% and 0.3% lower hit rate when analysed to the existing methods like MDMC-GAN, DMR-HMA-LSTM-AE and GDMB-CGAN respectively. In upcoming work, we are going to focus on a few important aspects. First, our goal is to increase the effectiveness of the suggested framework by integrating additional constraints aimed at mitigating drift when generating lengthy sequences. Secondly, our plan involves expanding the dataset to encompass dances involving multiple performers, thereby enriching the learning process with more comprehensive information. Lastly, we intend to pioneer novel methodologies for achieving stronger correlations between music and dance. This may involve endeavourslike choreographing dances inspired by music and synthesizing music based on dance sequences.

REFERENCE

[1] Aristidou, A., Yiannakidis, A., Aberman, K., Cohen-Or, D., Shamir, A., &Chrysanthou, Y. (2022). Rhythm is a dancer: Music-driven motion synthesis with global structure. *IEEE transactions on visualization and computer graphics*.

[2] Mendoza, I., Will-Cole, A., &Lamberson, L. (2023). The Mechanics of Dance: Using Parametric Equations as Inspiration for Dance Choreography. *Creativity Research Journal*, *35*(1), 33-43.

[3] Lin, S., Zukerman, M., & Yan, H. (2024). Music-Driven Choreography Based on Music Feature Clusters and Dynamic Programming. *IEEE Transactions on Multimedia*.

[4] Novosadova, A., Luchenko, O., Filimonova-Zlatohurska, Y., Pakholchak, Y., &Kulyk, D. (2024). Development of artistic and imaginative thinking of the future choreography teacher in the process of professional training. *Research in Dance Education*, 1-13.

[5] Chakrabarty, S. (2023). Design and Implementation of a Computational Model for BharataNatyam Choreography. *In Computer Assisted Music and Dramatics: Possibilities and Challenges* (pp. 165-175). Singapore: Springer Nature Singapore.

[6] Whitley, A., Kirchhof, S., &Strutt, D. (2023). Digital Dance Studio VR (DDS-VR) An innovative user-focused immersive software application for digital choreographic composition, planning, teaching, learning, and rehearsal. *In ACM SIGGRAPH 2023 Immersive Pavilion* (pp. 1-2).

[7] Miao, Y. (2023). An Introduction to the Relationship between Dance Choreographers and Dancers: Taking the dance work" Looking at the Sea Tide" as an Example. *Highlights in Art and Design*, *4*(2), 145-148.

[8] Yadaei, A., Azadehfar, M. R., & Alizadehashrafi, B. (2022). Creation of Melodic and Rhythmic Patterns Based on Cultural Heritage: Sonification of Azerbaijani Dance Movements. *Journal of Sound and Music in Games*, *3*(2-3), 1-27.

[9] Li, R., Zhang, Y., Zhang, Y., Zhang, H., Guo, J., Zhang, Y., & Li, X. (2024). Lodge: A Coarse to Fine Diffusion Network for Long Dance Generation Guided by the Characteristic Dance Primitives. arXiv preprint arXiv:2403.10518.

[10] Yang, L. (2022). Influence of Human–Computer Interaction-Based Intelligent Dancing Robot and Psychological Construct on Choreography. *Frontiers in Neurorobotics, 16*, 819550.

[11] Wang, L. J. (2022). Research on the Application of Hybrid Density Network Combined with Gaussian Model in Computer Music Choreography. *Journal of Sensors, 2022*.

[12] Primadiana, F., &Astuti, K. S. (2023). Creative Process of Creating MuliButapis Dance by Maria Reni Wulandari at the Children's Choreography Test II in 2022 at Yogyakarta State Unversity. *International Journal of Multicultural and Multireligious Understanding*, *10*(5), 332-339.

[13] Németh, B., Lelkó, A., Antal, Z., & Csaba, A. (2022). Collision-free trajectory design for dance choreography of virtual drones in hierarchical structure. *IFAC-PapersOnLine*, *55*(6), 502-507.

[14] Bench, H., Candelario, R., Perillo, J. L., & Rosa, C. F. (2023). Choreographic Analysis as Dance Studies Methodology: Cases, Expansions, and Critiques. *In Dance Research Methodologies* (pp. 113-133). Routledge.

[15] Arjona, K. (2023). Making it Stick: Exploring Teaching Tools for Choreography Retention in High School Musica*l*s.

[16] Akçalı, E., Springer, A., Burley, X., Ballin, C. D., & Colon, E. (2024). From Value Stream Mapping to Choreographic Tension Mapping: Preliminary Results and Insights from an Exploratory Study. *Journal of Dance Education*, 1-14.

[17] Parrott, N. D. (2022). Make Up! An Exploration of Choreographic Methods: Transforming Sailor Moon Into Concert Dance.

[18] Calissendorff, M., & Jaresand, S. (2023). 'Listening is done by using all the senses'a study of how dancers create, communicate and perceive a choreographed dance. *Research in Dance Education,* 1-13.

[19] Irani, Z. (2023). Research Proposal: Exploring Phenomenological Change and Improving Attunement Through the Use of Dance/Rhythmic Movements. *International Journal of Transactional Analysis Research & Practice*, *14*(1), 55-63.

[20] Jiang, D. (2022). Matching Model of Dance Movements and Music Rhythm Features Using Human Posture Estimation. *Computational Intelligence and Neuroscience*, 2022.

[21] Sun, G., Wong, Y., Cheng, Z., Kankanhalli, M. S., Geng, W., & Li, X. (2020). Deepdance: music-to-dance motion choreography with adversarial learning. *IEEE Transactions on Multimedia, 23,* 497-509.

[22] Gong, W., & Yu, Q. (2021). A deep music recommendation method based on human motion analysis. *IEEE Access*, *9*, 26290-26300.
[23] Huang, Y. F., & Liu, W. D. (2021). Choreography cGAN: generating dances with music beats using conditional generative adversarial networks. *Neural Computing and Applications*, *33*, 9817-9833.
[24] Kritsis, K., Gkiokas, A., Pikrakis, A., &Katsouros, V. (2022). Danceconv: Dance motion generation with convolutional networks. *IEEE Access, 10*, 44982-45000.
[25] Kim, D., Kim, T., Lee, I., & Lee, S. (2024). Kinematic Diversity and Rhythmic Alignment in Choreographic Quality Transformers for Dance Quality Assessment. *IEEE Transactions on Circuits and Systems for Video Technology*.
[26] Cui, C., Li, J., Du, D., Wang, H., Tu, P., & Cao, T. (2021). The method of dance movement segmentation and labanotation generation based on rhythm. *IEEE Access, 9*, 31213-31224.
[27] Kong, S. (2024). A deep learning model of dance generation for young children based on music rhythm and beat. *Concurrency and Computation: Practice and Experience*, e8053.
[28] https://google.github.io/aistplusplus_dataset/factsfigures.html
[29] Zhan, M., Zhao, C., Qin, K., Huang, P., Fang, M., & Zhao, C. (2023). Subaperture keystone transform matched filtering algorithm and its application for air moving target detection in an SBEWR system. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16*, 2262-2274.
[30] Shang, X., Li, G., Jiang, Z., Zhang, S., Ding, N., & Liu, J. (2024). Holistic Dynamic Frequency Transformer for image fusion and exposure correction. *Information Fusion, 102*, 102073.
[31] Zhou, Y., Zheng, Q., Wang, Y., Yan, W., Shi, P., & Zhu, J. (2024). MCoCo: Multi-level Consistency Collaborative multi-view clustering. *Expert Systems with Applications, 238,* 121976.
[32] Li, Y., Shi, T., Zhang, Y., & Ma, J. (2023). SPGAN-DA: Semantic-preserved generative adversarial network for domain adaptive remote sensing image semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*.
[33] Zareian, L., Rahebi, J., &Shayegan, M. J. (2024). Bitterling fish optimization (BFO) algorithm. *Multimedia Tools and Applications*, 1-34.