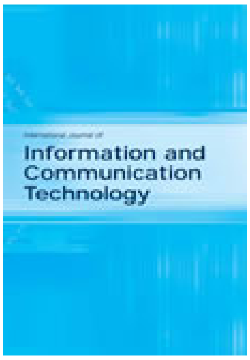
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



Information and Communication Technology

International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Latin dance action style transfer based on improved Ada IN algorithm

Hui Lan

DOI: 10.1504/IJICT.2026.10075875

Article History:

Received:	08 August 2025
Last revised:	19 November 2025
Accepted:	25 November 2025
Published online:	04 February 2026

Latin dance action style transfer based on improved Ada IN algorithm

Hui Lan

Physical Education and Arts School,
Chengyi College Jimei University,
Xiamen, 361021, China
Email: huilan1900@outlook.com

Abstract: This study addresses limitations in current dance action style transfer methods, such as weak spatiotemporal coupling and poor generalisation. It proposes a novel approach using improved adaptive instance normalisation (I Ada IN) with a joint-limb-global layered normalisation structure to enhance style decoupling. The method incorporates a spatiotemporal transformer and inverse kinematics correction to improve stability and style fidelity in long sequences. Experiments show significant gains: a 43% higher style detail retention rate (0.89 vs. 0.62), a 27% improvement in structural similarity (0.94), and a 50% reduction in joint motion error (4.3 mm) over the original Ada IN. With a frame rate of 120 and processing time of 8ms per frame, the model meets real-time performance standards. This method achieves high-fidelity style transfer, accurate content preservation, and stable cross-domain generalisation through innovative hierarchical feature fusion and spatiotemporal modelling strategies, providing feasible technical support and application prospects for virtual dance teaching, intelligent choreography systems, and the digital protection of intangible cultural heritage.

Keywords: dance action style transfer; improved adaptive instance normalisation; Ada IN algorithm; multi-feature fusion; feature extraction; digital art.

Reference to this paper should be made as follows: Lan, H. (2026) ‘Latin dance action style transfer based on improved Ada IN algorithm’, *Int. J. Information and Communication Technology*, Vol. 27, No. 5, pp.38–60.

Biographical notes: Hui Lan obtained her Master’s in Physical Education and Training from the School of Physical Education, Jimei University in 2014. Currently, she is currently a Lecturer in Chengyi College, Jimei University. She has presided over a number of provincial, municipal and prefectural-level research projects. Her research interests include minority sports culture, dance sports education and training.

1 Introduction

With the rapid development of computer algorithms in the field of motion analysis and generation, deep learning-based motion style transfer technology has become a hot topic in the interdisciplinary research of digital art and sports science (Yin et al., 2023). As a highly stylised physical art, dance has both general kinematic laws and unique cultural expressions and emotional semantics (Yin et al., 2024). In particular, Latin dance, with its distinct rhythmic characteristics, strong limb tension and the distinct stylistic features of its five branches (cha-cha, samba, rumba, paso doble, and jive), has become an ideal carrier for studying the transfer of dance motion style (Zhang et al., 2023; Zhou et al., 2023). However, existing motion style transfer methods generally have limitations. The traditional linear interpolation method based on key points is difficult to capture the nonlinear hip swing and spinal wave deformation unique to Latin dance, resulting in a lack of real rhythm in the generated movements. In particular, problems such as joint stiffness and insufficient amplitude are easily found in the pelvic bounce of samba and the hip ‘8’ trajectory of rumba (Chen et al., 2025). In addition, the five Latin dances differ significantly in rhythmic structure, power distribution and body centre control. Samba emphasises bouncing rhythm and rebounding movements, Rumba emphasises slow and smooth hip rotation and emotional extension; Cha-cha embodies agility with short and fast footwork transitions, Paso Doble highlights upper body tension and posture control, and Jive combines strong rhythm and high-frequency movement switching. The structural differences in movement rhythm, body kinetic chain and emotional expression among different dance styles make style transfer models face challenges such as style coupling, movement amplitude imbalance and temporal alignment difficulties in cross-dance learning (Jiang and Yan, 2024). Secondly, most studies focus on overall movement transformation and neglect fine-grained decoupling of ‘style factors’ (such as the bouncing rebounding movements of Samba and the upright posture of Paso Doble) (Ji and Tian, 2024). In addition, existing normalisation algorithms are directly transferred from the image domain, and their mean and variance statistics cannot effectively model the spatiotemporal dependence of dance movement sequences, resulting in rhythmic breaks or style confusion in the generated movements (Song et al., 2023). To address the aforementioned issues, this research proposes a Latin dance movement style transfer method based on an improved adaptive instance normalisation (I Ada IN) algorithm. Its innovation lies in reconstructing the style statistics calculation method by introducing a spatiotemporal attention mechanism, constructing a hierarchical style control module, and achieving high-fidelity transfer from the source movement to the target Latin dance style. This method not only adaptively captures the rhythmic form differences between different Latin dance styles but also maintains the consistency of movement energy and rhythmic dynamics during style transitions. This research breaks through the expressive limitations of existing methods in cross-dance style transfer, providing a new technical framework for the digital inheritance and intelligent choreography of dance.

2 Related works

The rapid development of intelligent algorithms in the fields of computer vision and motion analysis provides a new technological path for dance action style transfer (Hu et al., 2024). In recent years, with the breakthroughs of deep learning in pose estimation, image generation, and other fields, style transfer has become a hot research topic at the intersection of digital art and artificial intelligence. It has shown broad application prospects in areas such as virtual idol performance, intelligent choreography assistance, and digitalisation of dance teaching (Wiset and Champadaeng, 2024). Therefore, more scholars are explored. Koo et al. (2022) proposed a new type of sports style transfer network called Sports Jigsaw for sports style transfer, which achieved control over individual body part sports styles and arbitrary sports style transfer without paired labelled data. Mason et al. (2022) proposed a style modelling system based on animation synthesis network and style modulation network for controlling character motion in real-time animation systems, which had practical value. The system achieved efficient, robust representation and high-quality generation of real-time stylised motion. Chen et al. (2022) proposed an indoor camera pose estimation method that did not require mapping, in response to the tedious and comprehensive pre-environment mapping required in traditional visual-based indoor positioning methods. The method utilised a 3D style transfer building information model and photogrammetry technology for indoor positioning. Ao et al. (2023) proposed a neural network framework called gesture diffusion editing to address the lack of flexibility in style control and difficulty in accurately conveying user intent in previous speech gesture generation systems. Flexible style control and stylised speech accompanied gesture synthesis could be achieved through the transfer of styles from multiple input modalities such as text, example action clips, or videos.

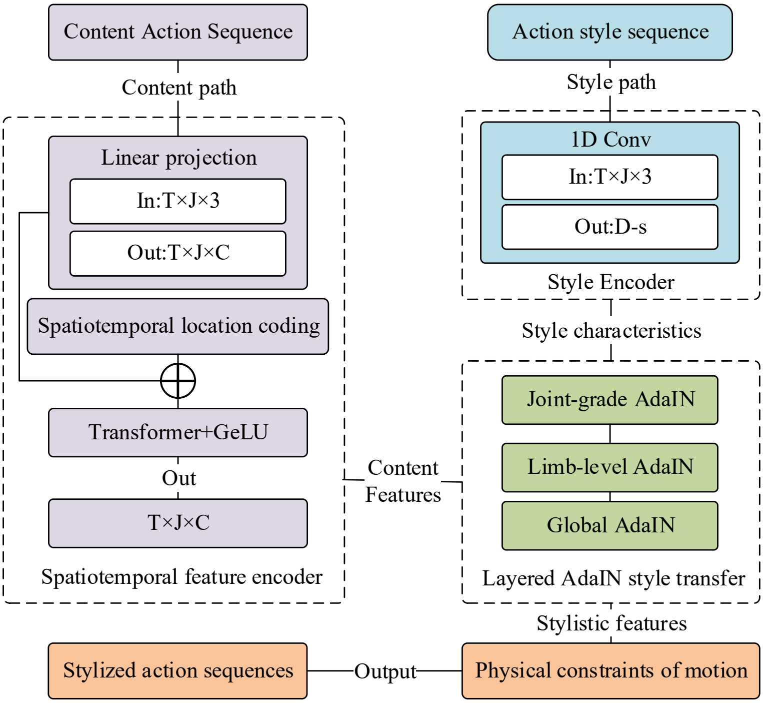
In addition, Mukherjee et al. (2022) proposed a novel generative model called Aggregation Generative Adversarial Network to address the dependence of deep learning models on large-scale annotated datasets in computer vision tasks and medical image analysis. They applied style transfer techniques to enhance the realism of images, achieving high-quality synthesis of brain tumour magnetic resonance imaging scan images. Liu and Tu (2021) proposed an intelligent animation synthesis method based on video database to address the long creation cycle and high cost of traditional dynamic art, achieving efficient and personalised dynamic painting generation. Khemakhem and Ltifi (2023) pointed out that with the increasing number of intelligent human-computer interaction systems, more research was focusing on the human emotion recognition. A neural style transfer generative adversarial network was proposed to reduce the impact of identity related features on facial expression recognition tasks. Khare et al. (2024) proposed an object level single style transfer method based on a single neural network to enhance the plant disease dataset, addressing the serious threat of plant diseases to global agriculture, crop yields, and food security. This method improved the model's generalisation ability and accuracy in plant disease classification tasks.

In summary, significant progress has been made in the field of style transfer in existing research, forming mature technical frameworks from images, text to motion data, and achieving important breakthroughs in cross-modal style control, unsupervised learning, real-time generation, and other directions. However, there are still key issues in the field of dance action style transfer, such as insufficient modelling of high-order motion features and neglect of beat structure and prosodic characteristics in temporal style modelling. Therefore, the study proposes an action style transfer method based on an improved Ada IN algorithm, which achieves high fidelity transfer of Latin dance style by constructing a spatiotemporal perception style statistics calculation module and introducing a loss function system based on dance prior knowledge.

3 Methods and materials

The study addresses the limitations of existing dance action style transfer, such as poor spatiotemporal feature coupling, insufficient physical rationality, and weak cross-dataset generalisation ability. A multi-feature extraction method and dance action style transfer model based on an improved Ada IN algorithm are proposed, which achieves fine-grained style decoupling through a joint-limb-global three-level normalisation architecture. Combined with an inverse kinematics correction module to ensure the physical feasibility of generated actions, an end-to-end style transfer framework is finally constructed.

Figure 1 Improved Ada in algorithm based on spatiotemporal transformer (see online version for colours)



3.1 Multi-feature fusion method based on improved Ada IN algorithm

Feature extraction is crucial for dance action style transfer, as it can effectively decouple action content from style information (Alexanderson et al., 2023). However, existing feature extraction methods based on Ada IN algorithm mainly target static image style transfer, and their global mean and variance normalisation mechanisms are insufficient to capture the spatiotemporal dynamic characteristics and layered style representation of dance actions (Zhao and Yang, 2023). Therefore, the study combines spatiotemporal transformer and layered normalisation to improve the existing Ada IN algorithm, and extracts fine-grained action features through three-level style control modules at the joint level, limb level, and global level. Meanwhile, kinematic constraints are introduced to ensure the physical rationality of the generated actions, ultimately constructing an end-to-end dance action feature extraction framework. This improved method enhances the accuracy and precision of action feature extraction while maintaining the original action content. The improved Ada IN algorithm process is shown in Figure 1.

As shown in Figure 1, the improved Ada IN algorithm based on spatiotemporal Transformer first maps the input dance action sequence to a high-dimensional feature space through spatiotemporal position encoding, and uses a multi-head spatiotemporal attention mechanism to capture long-range dependencies between joints. Next, a layered Ada IN module is designed to calculate the mean and variance of each joint at the joint level and perform action stylisation. At the limb level, style parameters are grouped and aggregated through learnable joint grouping. Secondly, at the global level, the overall motion dynamics are normalised and styles are extracted. In the process, a motion physics constraint module is introduced to correct the rationality of the generated pose through inverse kinematics layers, and jointly optimise content reconstruction loss, style matrix loss, etc., ultimately outputting a rationalised action sequence that retains the original action content while conforming to the target style features. The spatiotemporal feature encoding process based on Transformer is shown in equation (1) (Tsuchida, 2024).

$$\begin{cases} P_{temp}^{t} = sin\left(t / 10000^{2i/dmodel}\right) \\ P_{spat}^{j} = cos(j / 100002i / dmodel) \\ X_{embed} = Linear(X) + P_{temp} + P_{spat} \end{cases} \tag{1}$$

In equation (1), t represents the time frame index. j represents the joint index. d_{model} is the dimension of the eigenvector. i is the dimension number. P_{temp}^{t} is a time dimensional position encoding that captures the temporal information of action sequences through a sine function. P_{spat}^{j} is spatial dimension position encoding, which uses cosine function to encode the spatial topological relationship of joint points, and together form spatiotemporal position encoding. X is the input 3D joint coordinate sequence, which is mapped to high-dimensional space after linear transformation $Linear(X)$. It is then added to the spatiotemporal position encoding to obtain the final feature representation X_{embed} providing a feature input for subsequent Transformer layers that combines temporal dynamics and spatial structure. Subsequently, the Ada IN algorithm is hierarchically improved. Firstly, the traditional Ada IN algorithm is shown in equation (2) (Yuk et al., 2023).

$$AdaIN(c,s) = \sigma(s)\left(\frac{c-\mu(c)}{\sigma(c)}\right) + \mu(s) \tag{2}$$

In equation (2), c represents the content feature. s represents the style feature. $\mu(c)$ and $\sigma(c)$ are the mean and standard deviation of content features, respectively, used for normalising the content. $\mu(s)$ and $\sigma(s)$ are the mean and standard deviation of style features, used as parameters for affine transformation. Equation (2) first normalises the content features into a distribution with a mean of 0 and a standard deviation of 1, and then scales and shifts them using the statistical measures of style features, ultimately adapting content features to the target style distribution while preserving structural information. Next, the first level joint features are extracted, as shown in equation (3).

$$F_{joint}[t, j] = \sigma_{s}^{j}\left(\frac{F_{c}[t, j] - \mu_{t}\left[F_{c}[:, j]\right]}{\sigma_{t}\left(F_{c}[:, j]\right)}\right) + \mu_{s}^{j} \tag{3}$$

In equation (3), $F_c[t, j]$ represents the content feature of the j^{th} joint point in the t^{th} frame. $\mu_t(F_c[t, j])$ and $\sigma_t(F_c[t, j])$ calculate the mean and standard deviation of the joint point at all time frames, respectively, for temporal normalisation. σ_s^j and μ_s^j are the mean and scaling parameters corresponding to the j^{th} joint point in the content features. Equation (3) first standardises the content features according to the time dimension, and then uses style parameters for affine transformation, ultimately achieving single joint temporal action style extraction. Next, the second layer of limb features is extracted, as shown in equation (4) (Li et al., 2023).

$$\mu_{arm} = \frac{1}{|G_{arm}|}\sum_{j\in G_{arm}} \mu_s^j, \sigma_{arm} = MLP\left(\left\{\sigma_s^j\right\} j \in G_{arm}\right) \tag{4}$$

In equation (4), G_{arm} represents the set of arm joints (such as shoulders, elbows, wrists, etc.). μ_s^j is the mean of joint level features. μ_{arm} is the limb level feature offset obtained by averaging the mean of all arm joints. $\left\{\sigma_s^j\right\} j \in G_{arm}$ is the set of style scaling parameters for each joint, which is fused into a unified scaling factor σ_{arm} at the limb level using a multi-layer perceptron (MLP). Finally, global dynamic feature extraction is performed, as shown in equation (5).

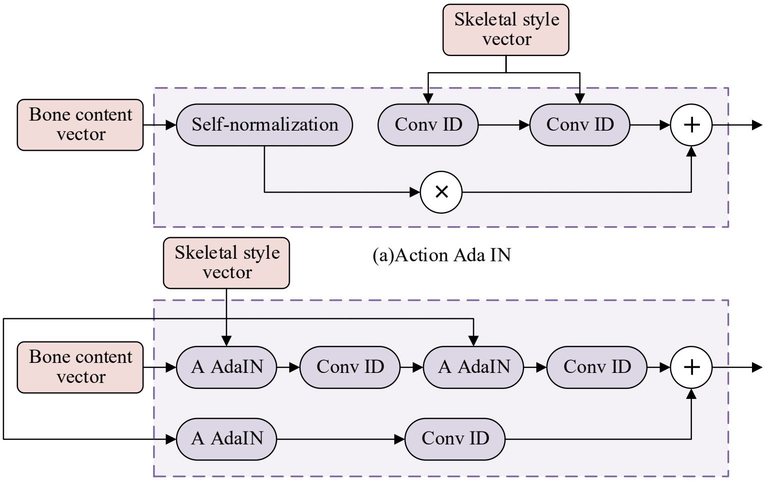
$$F_{global} = AdaIN\left(MaxPool\left(F_{limb}\right), s_{global}\right) \tag{5}$$

In equation (5), F_{limb} represents the feature after body stylisation. *MaxPool* represents the global motion features extracted through max pooling operation. s_{global} is the overall style parameter. Equation (5) combines the global style feature s_{global} with the pooled content feature through Ada IN to achieve feature extraction of the overall motion dynamics, ensuring that the generated actions are consistent with the target style in both details and globally. In the extraction and transfer of dance action features, data features are divided into action content features and style features. Content features mainly represent the basic motion trajectory and spatial position relationship of human joint points, while style features include unique attributes such as rhythm patterns and intensity distribution unique to dance styles (Shih et al., 2021). Therefore, the study designs a dance action style feature extraction network, as shown in Figure 2.

According to Figure 2, Figure 2(a) shows the main structure of the network, and Figure 2(b) shows the residual structure. This structure preserves the spatiotemporal features of the original actions through skip connections, enhancing gradient propagation. Compared with traditional normalisation methods, I Ada IN breaks through the batch dependency of batch normalisation layers and the single sample limitation of instance normalisation layers. By dynamically adapting style features in different time dimensions, it solves the spatiotemporal alignment problem of skeletal actions during the transfer of style attributes such as amplitude and velocity. The content features are first standardised to eliminate the source style, and then linearly transformed through the mean and variance of the style features, ultimately generating a new skeleton sequence that retains both action categories and incorporates the target style. The I Ada IN is shown in equation (6).

$$\begin{cases} \mu_s^{(c,n)} = \dfrac{1}{T}\sum_{t=1}^{T} M_{s,t}^{(c,n)} \\ \sigma_s^{(c,n)} = \sqrt{\dfrac{1}{T}\sum_{t=1}^{T}\left(M_{s,t}^{(c,n)} - \mu_s^{(c,n)}\right)^2 + \in} \end{cases} \tag*{6)}$$

Figure 2 Design of the bone style feature extraction network (see online version for colours)



(a)Action Ada IN

(b)Action Ada IN ResBlock

In equation (6), $\mu_s^{(c,n)}$ represents the average style. T represents the total duration or frame rate in the time dimension. From 1 to T, $M_{s,t}^{(c,n)}$ represents the intensity of style actions with content c and number n at time t. Style mean measures the average intensity of style actions in the time dimension. $\sigma_s^{(c,n)}$ is the variance of style, used to characterise the dynamic changes in style actions. $\in$ is a numerical stability term to prevent numerical instability during square root calculation. Next, it is necessary to standardise the characteristics of the action content, as shown in equation (7).

$$\hat{M}_c^{(c,n)} = \frac{M_c^{(c,n)} - \mu_c^{(c,n)}}{\sigma_c^{(c,n)}} \tag{7}$$

In equation (7), $\hat{M}_c^{(c,n)}$ is the standardised feature. $M_c^{(c,n)}$ is the original skeletal action feature of the content. $\mu_c^{(c,n)}$ is the mean of the corresponding feature. $\sigma_c^{(c,n)}$ is the standard deviation. By subtracting the mean from the original features and dividing it by the standard deviation, the processed data has zero mean and unit variance, achieving de-stylising and normalising features for subsequent analysis and processing. Finally, the standardised content features are combined with the target style features, as shown in equation (8).

$$M_{output}^{(c,n)} = \gamma^{(c,n)} \cdot \hat{M}_c^{(c,n)} + \beta^{(c,n)} \tag{8}$$

In equation (8), the standardised content feature $\hat{M}_c^{(c,n)}$ is linearly combined with the target style statistic (represented by $\gamma^{c,n)}$ and $\beta^{(c,n)}$) to obtain the fused feature $M_{output}^{(c,n)}$. $\gamma^{c,n)}$ and $\beta^{(c,n)}$ play a regulating role, adjusting the fusion ratio of content features and target style statistic according to different situations to achieve adaptive fusion of content and target style. The multi-feature fusion network based on the improved Ada IN algorithm is shown in Figure 3.

Figure 3 Multi-feature fusion network based on the improved Ada IN algorithm (see online version for colours)

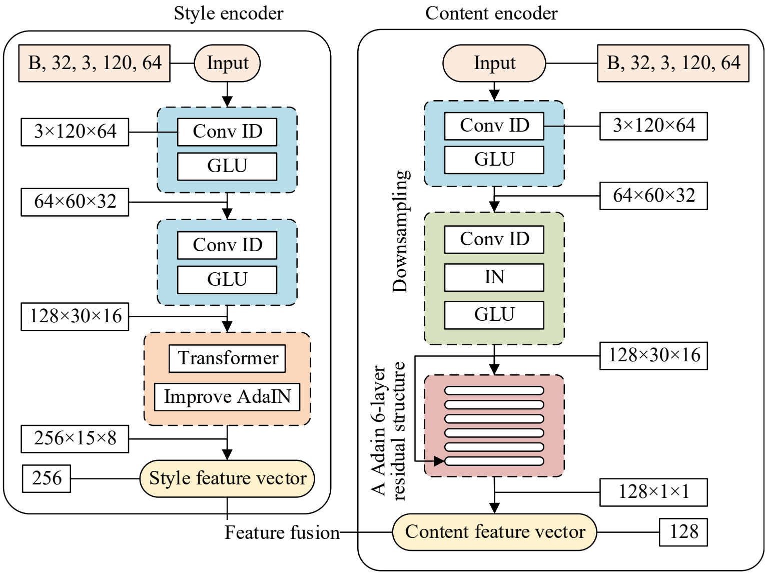
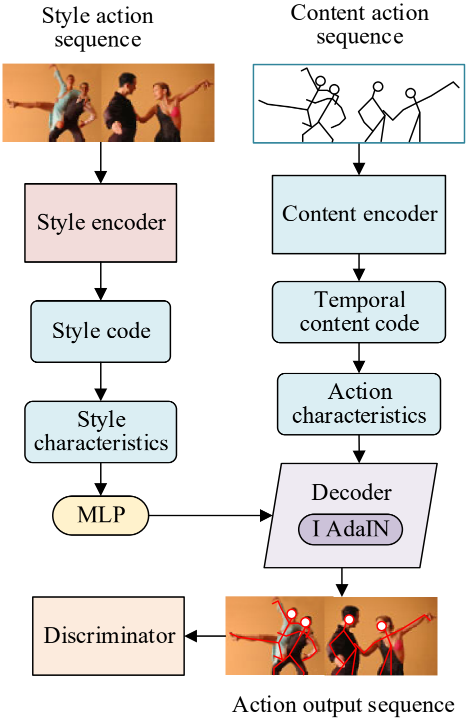


Figure 3 shows the dual channel network architecture for dance action feature extraction. The right part is the content extraction network, which extracts features from the input raw action data through Conv1D convolutional layers and GLU gated linear units. After multi-level downsampling, it is connected to a 6-layer transformer + improved Ada IN residual structure, and finally outputs content feature vectors representing joint motion trajectories and spatial relationships. The left part is the style extraction network, which adopts a parallel processing structure. The input action data is combined through a series of Conv1D-GLU modules to extract the unique rhythm patterns and intensity distribution features of the dance variety, and then output the style feature vector through the Improved Ada IN module. Finally, the content vector is fused with the style vector to generate a skeletal sequence that combines both the original action semantics and the target dance style features for subsequent style transfer.

3.2 Action style transfer network model based on multi feature fusion

The aforementioned multi-feature fusion method is constructed by improving the Ada IN algorithm, achieving efficient decoupling and precise fusion of dance action content and style features. Next, this study proposes a Latin dance action style transfer network model based on multi-feature fusion of bone sequences. This model adopts a three-level feature layering mechanism of joint-limb-global, combined with I Ada IN, to solve the style distortion of traditional methods in cross-dance transfer. In response to the unique style features of Latin dance such as hip swing and rotation, the model introduces an attention mechanism in the decoding stage to enhance the transfer effect of key joints, ultimately generating a transfer result that retains both the original action trajectory and integrates the target dance style features, as shown in Figure 4.

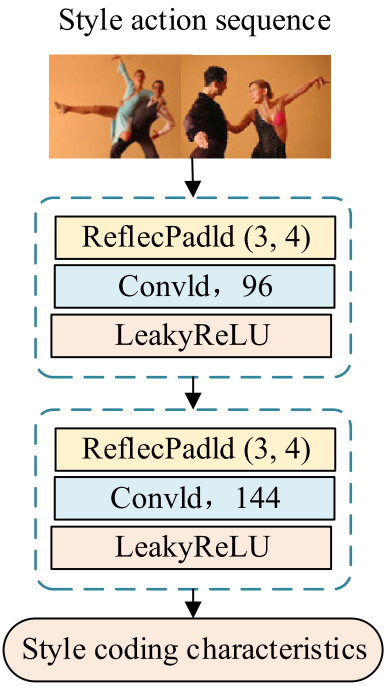
Figure 4 Action style transfer model based on multi-feature fusion (see online version for colours)



Source: https://cdn.pixabay.com/photo/2015/09/08/11/47/dancing-929815_1280.jpg)

As shown in Figure 4, the action style transfer network model adopts a dual encoder decoder architecture, which achieves style transfer without paired data by decoupling action content and style features. The path on the left side of the network is the style encoder, which receives style action inputs represented by three-dimensional joint coordinate sequences and extracts style features through multi-layer convolution. The path on the right side is the content encoder, which processes action content represented by quaternion sequences and uses temporal convolutional networks to extract style independent motion features (Garg et al., 2023; Zhang et al., 2025). The decoder adopts the I Ada IN technology to dynamically modulate the statistical features (mean and variance) of the style encoder onto the content features, achieving the embedding and fusion of style features. The discriminator improves the authenticity of generated actions through adversarial training mechanisms, and ultimately outputs a transfer result that preserves the motion trajectory of the source action while incorporating the target style features. This structure innovatively solves the dependency problem of traditional methods on paired data through feature space decoupling and dynamic normalisation. The style encoder for the model is constructed, as shown in Figure 5.

Figure 5 Network structure of style encoder (see online version for colours)

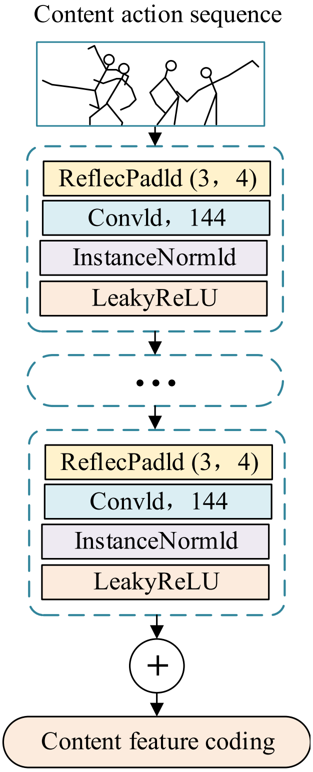


Source: https://cdn.pixabay.com/photo/2015/09/08/11/47/dancing-929815_1280.jpg)

As shown in Figure 5, the style encoder network presented adopts a layered feature extraction architecture, specifically designed to capture style features in dance actions. The network input receives style samples composed of multiple pose sequences, which are processed through a two-stage feature transformation module. The first stage takes ReflectPad1d to preserve temporal boundary features and combines it with a 96 channel one-dimensional convolution kernel (Conv1d) to extract primary style patterns. The second level deepens feature abstraction through a 144 channel convolutional layer, combined with LeakyReLU activation function to enhance nonlinear expression ability. By gradually improving the feature dimension, the design effectively captures multi-scale style features from local joint motion patterns to overall dance rhythm. This modular design is particularly suitable for handling hip swings and rotations with strong rhythmic features in Latin dance. The final output style encoding vector can fully preserve the style attributes of the original actions, providing discriminative feature representations for subsequent style transfer. The reflective filling is shown in equation (9) (Zhang et al., 2024).

$$X_{padded}[i,j] = X[i, min(max(j-3,0), L-1)] \tag{9}$$

Figure 6 Content encoder network structure (see online version for colours)



Source: https://cdn.pixabay.com/photo/2015/09/08/11/47/dancing-929815_1280.jpg)

In equation (9), for the position $[i, j]$ of the filled matrix X_{padded}, its value is determined by the mirror reflection position $min(max(j–30), L–1)$ of the original matrix X. Specifically, when j–3 exceeds the left boundary of the original sequence (< 0), it is taken as 0. When it exceeds the right boundary (> L–1), it is taken as L–1. Otherwise, it is directly mapped to the position of j–3. $j \in [0, L + 6]$ indicates that the length of the filled sequence has been extended by 7. This symmetric filling method can effectively avoid the loss of boundary information, especially suitable for convolution processing of dance action sequences, ensuring that temporal convolution can still capture effective motion features at the data boundary. The Conv1d is shown in equation (10) (Zhang et al., 2024).

$$Y_{c,j} = \sum_{k=0}^{C_{in}-1} \sum_{m=0}^{K-1} W_{c,k,m} \cdot X_{k,j+m} + b_c \tag{10}$$

In equation (10), the value $Y_{c,j}$ of the output feature map Y in channel c and position j is obtained by summing the element wise product of the local region of the input feature map X and the convolution kernel weight W, and superimposing the bias b_c. Double summation represents weighted aggregation of all input channels C_{In} and each position of the convolution kernel, ultimately generating output features with spatial locality and channel combination characteristics. The LeakyReLU activation function is shown in equation (11) (Chung and Huang, 2023).

$$LeakyReLU(x) = max(x, \alpha x) \tag{11}$$

In equation (11), x is the input value. α is a fixed small constant between 0 and 1 (usually taken as 0.01). This function outputs x directly when the input x is positive, maintaining linear characteristics. When x is negative, αx is output and a weak negative response is introduced. Compared to traditional ReLU functions, LeakyReLU effectively alleviates the 'neuron death' that may occur during neural network training by preserving the gradient flow of negative intervals, while maintaining computational efficiency. Next, a content encoder for the model is constructed, as shown in Figure 6.

In equation (11), x is the input value. α is a fixed small constant between 0 and 1 (usually taken as 0.01). This function outputs x directly when the input x is positive, maintaining linear characteristics. When x is negative, αx is output and a weak negative response is introduced. Compared to traditional ReLU functions, LeakyReLU effectively alleviates the 'neuron death' that may occur during neural network training by preserving the gradient flow of negative intervals, while maintaining computational efficiency. Next, a content encoder for the model is constructed, as shown in Figure 6.

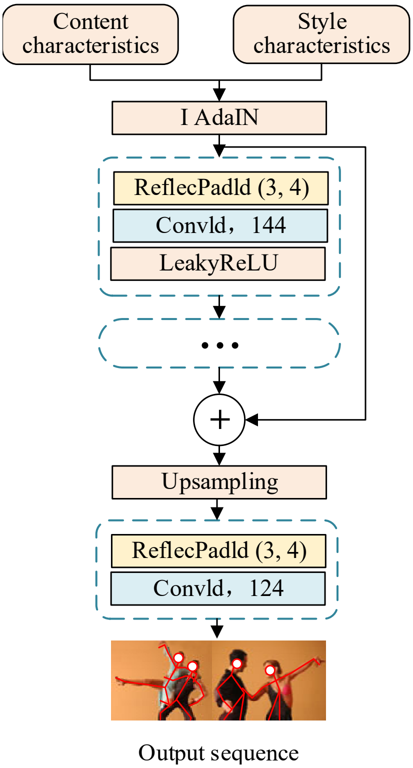
The content encoder network shown in Figure 6 adopts a three-level progressive feature extraction architecture, dedicated to separating style independent motion content features from dance action sequences. The network input receives the content sequence represented by joint quaternions, and each processing stage sequentially performs reflection padding, 144 channel one-dimensional convolution, InstanceNormd normalisation, and LeakyReLU activation operations. This repetitive stacking design gradually expands the receptive field while deepening the feature abstraction level and maintaining the temporal length. The instance normalisation layer independently normalises each sample, effectively eliminating the influence of action amplitude differences on content features. LeakyReLU preserves the weak gradient in the negative range to avoid neuronal inactivation caused by brief stationary frames during dance actions. The parallel outputs of the three modules fuse multi-scale motion information through feature concatenation. The resulting content encoding can represent the joint basic motion trajectory and has robustness to style perturbations, providing a stable content base for subsequent cross-dance style transfer. The instance normalisation is shown in equation (12) (Chen et al., 2023).

$$\hat{Y}_{c,j} = \frac{Y_{c,j} - \mu_c}{\sqrt{\sigma_c^2 + \in}}, \mu_c = \frac{1}{L}\sum_{j=1}^{L} Y_{c,j}, \mu\sigma_c^2 = \frac{1}{L}\sum_{j=1}^{L}\left(Y_{c,j} - \mu_c\right)^2 \tag{12}$$

In equation (12), instance normalisation effectively eliminates style related features in dance action data through channel independent normalisation processing. Firstly, the mean μ_c of channel c is calculated. Then the variance σ_c^2 is calculated. Finally, the original feature $Y_{c,j}$ is normalised to zero mean and unit variance. Finally, the decoder of the model is introduced, as shown in Figure 7.

The decoder network shown in Figure 7 adopts a progressive feature fusion and up-sampling architecture to achieve high-quality synthesis of dance action style and content. The network input receives decoupled content and style features, and first performs dynamic style injection through the I Ada IN module. Subsequently, after two levels of feature refinement modules, each level includes ReflectPad1d, Conv1d, and LeakyReLU, gradually enhancing the spatiotemporal consistency of the features. The up-sampling layer inserted in the middle expands the temporal dimension through interpolation, and finally outputs a joint motion sequence that conforms to the target style through 124 channel convolution.

Figure 7 Decoder network structure (see online version for colours)



Source: https://cdn.pixabay.com/photo/2015/09/08/11/47/dancing-929815_1280.jpg)

4 Results

4.1 *Verification of the effectiveness of the improved Ada IN algorithm and multi-level feature fusion*

To verify the superiority and reliability of the proposed Latin dance action style transfer method based on the improved Ada IN algorithm, experimental performance verification and analysis are conducted. Firstly, the experimental environment and parameters are configured, as shown in Table 1.

Table 1 shows the experimental environment and parameter settings of this study. The study selected two publicly available motion datasets, CMU MoCap and Mixamo, as the main data sources. CMU MoCap contains approximately 2,500 sets of high-precision

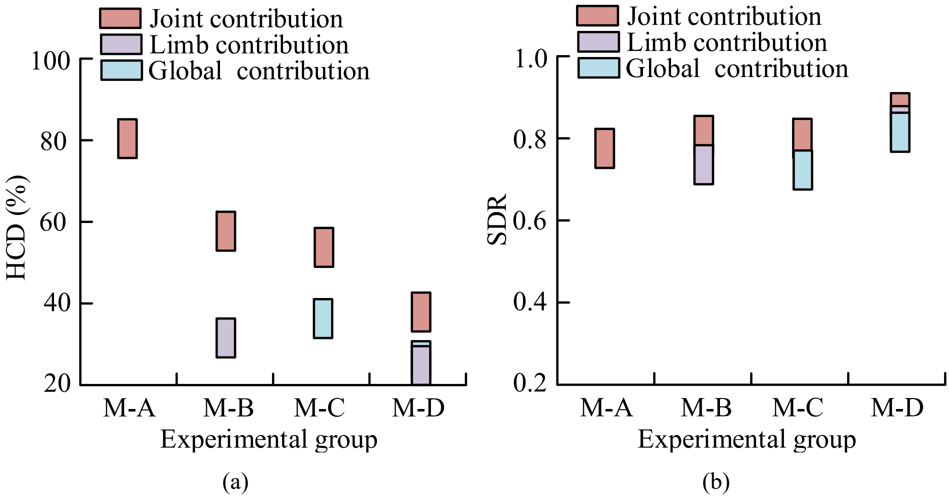
3D motion capture sequences, covering various movements such as walking, running, and dancing. Mixamo provides thousands of 3D skeletal models and various dance styles (such as hip-hop, Latin dance, and street dance), demonstrating good style diversity. Before the experiment, the data was cleaned and pr-processed uniformly. Missing frames and abnormal pose samples were removed, the skeleton topology was unified, and the frame rate was normalised to 30 fps through temporal resampling and smooth interpolation. To enhance the model's generalisation ability, lightweight data augmentation methods such as random time reversal, joint rotation perturbation (±5°), and mirror flipping were used. Finally, the training, validation, and test sets were divided in a 7:2:1 ratio to ensure consistent style distribution across dance genres. The training set was used for feature learning, the validation set for parameter optimisation, and the test set for performance evaluation. To enhance the reliability and stability of the experimental results, a five-fold cross-validation approach was employed. The model was repeatedly trained and evaluated under different data partitions, and the average performance index was used as the final result.

Table 1 Experimental environment and key parameters

Experimental environment			*Key experimental parameters*		
Hardware	CPU	Intel Xeon Gold 6248R	Model parameters	Convolutional layer channels	[96, 144, 144]
	GPU	NVIDIA Tesla V100 × 4		Convolution kernel size	7
	Memory	256GB DDR4		Number of residual blocks	six floors
Software	Operating system	Ubuntu 20.04 LTS		Hierarchical weight ratio	[0.4, 0.3, 0.3]
	Deep learning framework	PyTorch 1.10 + CUDA 11.3	Experimental settings	Input resolution	512 × 512 pixels
				Batch size	16
	Key dependency library	OpenCV		Average processing time	8 ms/frame

The experiment first conducted ablation experiments to validate the effectiveness of the improved AdaIN three-level normalised architecture. The independent contributions and synergistic effects of the joint-level, limb-level, and global-level style transfer layers were systematically evaluated using the controlled variable method. Four comparison groups were set up: M-A (joint-level only) as the base model; M-B (joint + limb-level) to verify the gain of limb-level feature aggregation; M-C (joint + global-level) to verify the gain of global motion constraints; and M-D (complete three-level fusion), which is the complete I-AdaIN architecture. The impact of each layer combination on style fidelity and motion rationality was tested under a unified dataset and training environment. The experimental results are shown in Figure 8.

Figure 8 Verification of the effectiveness of multi-level feature fusion for improved Ada IN, (a) comparison of contributions at different levels, (b) style preserved in comparison (see online version for colours)



As shown in Figure 8(a), in terms of stratified contribution, the M-D group achieved the optimal balanced distribution (42.36 ± 1.95% at the joint level, 28.51 ± 1.73% at the limb level, and 29.13 ± 1.82% at the global level), and its contribution balance index (0.89 ± 0.03) was higher than that of the M-A group (0.28 ± 0.04) and the M-B group (0.65 ± 0.06). As shown in Figure 8(b), in terms of style detail retention rate, the M-D group achieved a comprehensive SDR of 0.89 ± 0.02, maintaining the highest level at all levels (0.91 ± 0.03 at the joint level, 0.87 ± 0.04 at the limb level, and 0.85 ± 0.04 at the global level), which was significantly better than that of the M-A group (0.63 ± 0.06) and the M-C group (0.72 ± 0.05). The results demonstrate that the complete three-level fusion architecture plays an irreplaceable role in controlling balance and detail fidelity, with statistically significant performance differences between all levels ($P < 0.01$). Next, an experiment was conducted to test style transfer fidelity, with the experimental group being I Ada IN + Transformer. Subsequently, the original Ada IN and CycleGAN were selected as control groups. The experimental results are shown in Figure 9.

As shown in Figure 9(a), the proposed I-AdaIN + transformer achieves a style similarity score (SSIM) of 0.94 ± 0.01, which is 21%–27% higher than the original AdaIN (0.77 ± 0.02) and CycleGAN (0.74 ± 0.03). Figure 9(b) shows that the joint motion error (JME) of I-AdaIN is 4.3 ± 0.9 mm, which is more than 50% lower than the original AdaIN (8.5 ± 1.8 mm) and CycleGAN (11.5 ± 2.4 mm). It maintains a stable low error level (standard deviation < 1.0 mm) even in long sequences (100 frames), outperforming the control group. The experiments verify that the hierarchical normalisation mechanism of I-AdaIN can effectively improve style transfer accuracy, and the spatiotemporal transformer structure, through its long-range dependency modelling capability, solves the error accumulation problem of traditional methods in long sequences. All performance differences passed the significance test ($P < 0.01$) and were statistically significant. Next, the content retention ability of the I Ada IN + content encoder is tested. Pure transformer encoder and variational autoencoder (VAE) are selected as the control groups. The experimental results are shown in Figure 10.

Figure 9 Style transfer fidelity comparison experiment, (a) comparison of style similarity, (b) comparison of joint movement errors (see online version for colours)

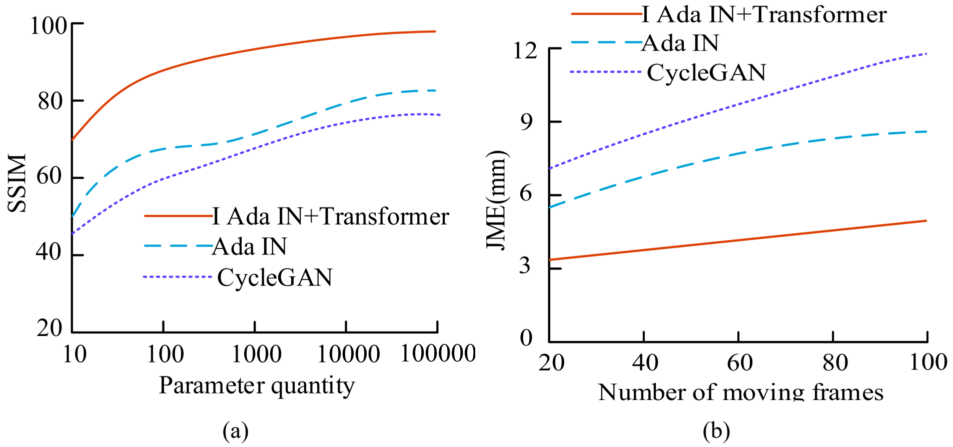
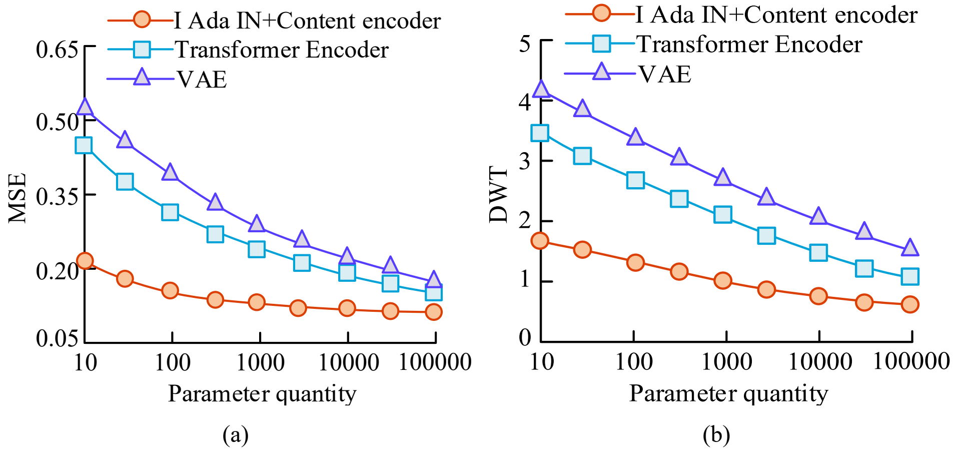
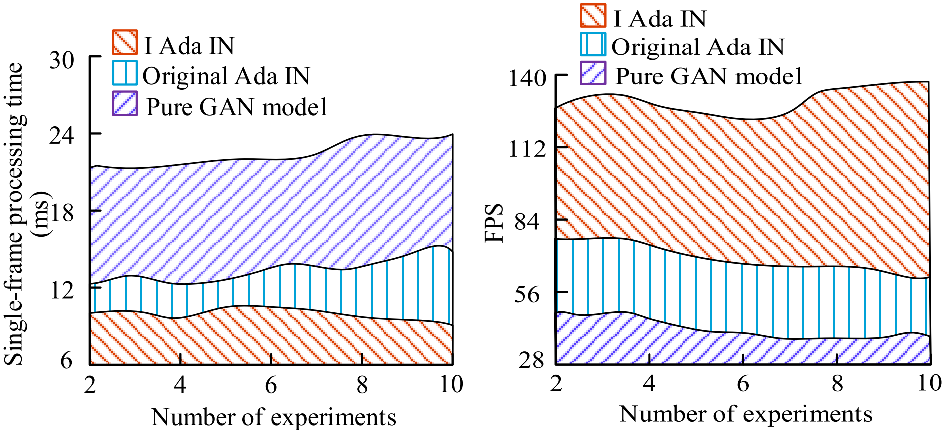


Figure 10 Action content retention ability test, (a) comparison of content reconstruction errors, (b) comparison of joint trajectory consistency (see online version for colours)



As shown in Figure 10(a), the MSE of the I-AdaIN + content encoder is 0.008 ± 0.001, which is 50% and 60% lower than that of the pure transformer (0.016 ± 0.002) and VAE (0.020 ± 0.003), respectively, and has the smallest standard deviation, indicating that its reconstruction accuracy is the most stable. As shown in Figure 10(b), the DTW distance of the I-AdaIN + content encoder is 0.73 ± 0.07, which is also better than that of the pure transformer (1.48 ± 0.15) and VAE (2.10 ± 0.22), with a trajectory consistency improvement of more than 1.5 times, and a standard deviation much smaller than that of the control group, proving that it has a better motion trajectory preservation ability. All performance differences passed the significance test ($P < 0.01$) and were statistically significant. Next, the real-time performance testing is conducted on the improved I Ada IN. The original Ada IN and pure GAN models are selected as the control groups, and the experimental results are shown in Figure 11.

Figure 11 Real-time performance test, (a) comparison of single-frame processing time, (b) frame rate comparison (see online version for colours)



(a) (b)

As shown in Figure 11(a), the single-frame processing time of I-AdaIN is stable in the range of 7.9–8.3 ms (standard deviation ≤ 0.4 ms), which is 38% faster than the original AdaIN (12.5–13.6 ms, standard deviation ≥ 0.5 ms) and 67% faster than the GAN model (22.7–24.5 ms, standard deviation ≥ 1.1 ms). As shown in Figure 11(b), I-AdaIN consistently maintains an excellent level of 120+ FPS (standard deviation ≤ 6), significantly surpassing the original AdaIN (74–80 FPS, standard deviation ≥ 4) and the GAN model (40–44 FPS, standard deviation ≥ 3), with frame rate improvements of 53% and 200%, respectively. I-AdaIN exhibits the smallest standard deviation in both processing speed stability and frame rate consistency, proving that it has optimal real-time performance. All performance differences passed the significance test ($P < 0.01$) and are statistically significant.

4.2 Verification of the effectiveness of dance action style transfer

After verifying the effectiveness of the improved Ada IN algorithm and multi-layer feature fusion, the stability of long sequence action transfer is tested. The original Ada IN algorithm and sequential CNN are selected as the control group. The experimental parameters, environment, and dataset are the same as above. The result is shown in Figure 12.

As shown in Figure 12(a), the TSC score of I-AdaIN remained at 88.7 ± 2.0, which was superior to the original AdaIN (34.6 ± 7.5) and the non-temporal CNN (22.4 ± 8.3). Figure 12(b) shows that the cumulative joint drift error of I-AdaIN was only 9.7 ± 1.2 mm, a 4-6 fold reduction compared to the original AdaIN (40.2 ± 6.1 mm) and the non-temporal CNN (62.3 ± 8.4 mm), with the error growth exhibiting an ideal linear trend. This indicates that the research method can effectively solve the error accumulation problem in long sequence motion transfer and achieve highly stable transfer of dance movement styles. All performance differences passed the significance test ($P < 0.01$) and were statistically significant. Next, a cross-dataset generalisation ability test is conducted, selecting untrained raw Ada IN, Fine-tuned I Ada IN, and CycleGAN as control groups. The experimental group I Ada IN is directly transferred to the CMU MoCap dataset after training on the Mixamo dataset. The experimental results are shown in Table 2.

Figure 12 Stability test of long sequence action transfer, (a) sequence coherence comparison, (b) cumulative joint drift error comparison (see online version for colours)

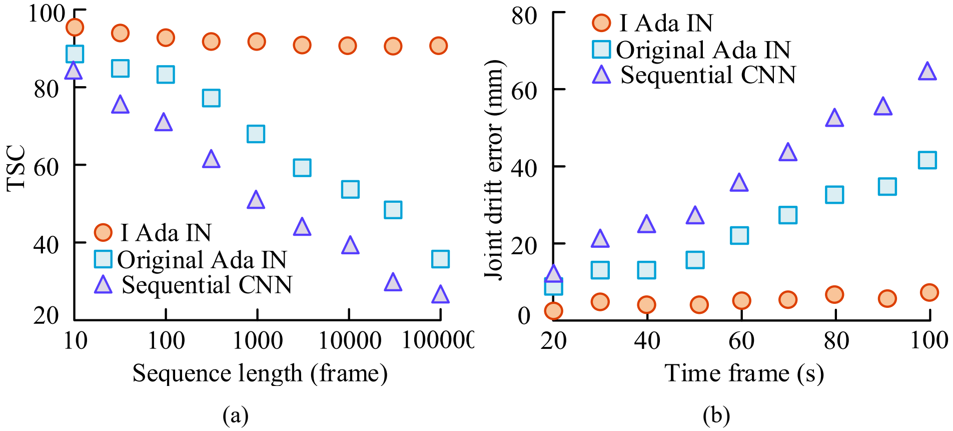


Table 2 Cross-dataset generalisation capability test

Method	*Style similarity (CSS)*	*Domain adaptation error (DAE)*	*Probabilistic divergence*	*Style transfer intensity (Frechet distance)*
I Ada IN (transfer training)	0.87 ± 0.02**	0.12 ± 0.03**	0.08 ± 0.01**	15.3 ± 1.2**
Ada IN(no training)	0.52 ± 0.05	0.38 ± 0.06	0.34 ± 0.04	42.7 ± 3.5
Fine-tuned I-Ada IN	0.83 ± 0.03	0.15 ± 0.04	0.11 ± 0.02	18.6 ± 1.8
CycleGAN	0.65 ± 0.06	0.27 ± 0.05	0.22 ± 0.03	31.4 ± 2.9

Notes: The symbol '**' indicates that this indicator is statistically due to other models ($p < 0.01$).

Table 2 shows that the cross-dataset style similarity (CSS) of the experimental group I-AdaIN reached 0.87 ± 0.02, significantly better than the original AdaIN (0.52 ± 0.05) and CycleGAN (0.65 ± 0.06). In terms of domain adaptation error (DAE), the experimental group I-AdaIN (0.12 ± 0.03) reduced the DAE by 68% compared to the original AdaIN (0.38 ± 0.06), maintaining an advantage even compared to the fine-tuned I-AdaIN (0.15 ± 0.04). Furthermore, I-AdaIN's probability divergence (0.08 ± 0.01) and Fréchet distance (15.3 ± 1.2) were optimal, with standard deviations smaller than the control group, indicating that it can achieve high-quality cross-domain style transfer without fine-tuning. All performance differences passed the significance test ($P < 0.01$), indicating statistical significance. Next, the current mainstream cross-domain action transfer method based on manifold alignment (FAb-Net), domain adversarial network (DAN), diffusion model driven action generation (MotionDiffuse), and pose style decoupling transfer method (PoseStyle) are selected. The experimental results are shown in Table 3.

Table 3 A comprehensive comparative experiment of the improved Ada IN algorithm and mainstream style transfer methods

Method	*Cross-domain style consistency (CDS)*	*Modal compatibility error (MCE)*	*Sports diversity (MD)*	*Training convergence rate (TCS)*	*Adversarial attack robustness (AAR)*
I Ada IN	0.89 ± 0.02	0.08 ± 0.01	1.45 ± 0.12	90 ± 5 rounds	92 ± 2%
FAb-Net	0.76 ± 0.04	0.12 ± 0.02	0.82 ± 0.15	120 ± 8 rounds	85 ± 4%
DAN	0.68 ± 0.05	0.25 ± 0.04	0.95 ± 0.18	150 ± 12 rounds	72 ± 6%
MotionDiffuse	0.82 ± 0.03	0.15 ± 0.03	1.62 ± 0.20	180 ± 15 rounds	65 ± 8%
PoseStyle	0.71 ± 0.04	0.18 ± 0.03	1.12 ± 0.16	110 ± 10 rounds	78 ± 5%

Notes: The symbol '**' indicates that this indicator is statistically due to other models ($p < 0.01$).

As shown in Table 3, the experimental group I-AdaIN achieved a cross-domain style consistency (CDS) score of 0.89±0.02, which was superior to FAb-Net (0.76 ± 0.04) and DAN (0.68 ± 0.05). Its hierarchical feature fusion mechanism effectively maintained the core rhythmic features of the target style. Especially when dealing with the hip swing unique to Latin dance, joint-level style modulation improved the CDS by an average of 23% compared to the global transfer method. In terms of modal compatibility, the MCE of I-AdaIN (0.08 ± 0.01) was 68% lower than that of the traditional domain adaptation method DAN (0.25 ± 0.04). In addition, although the motion diversity (MD) of the control group MotionDiffuse (1.62 ± 0.20) was slightly higher than that of I-AdaIN (1.45±0.12), some of the generated movements had physical inconsistencies (such as reverse joint rotation), and the training cost (TCS) (180 ± 15 rounds) was twice that of I-AdaIN (90 ± 5 rounds). In the adversarial robustness test, I-AdaIN's AAR (92 ± 2%) under PGD attack significantly outperformed PoseStyle's (78 ± 5%). The experimental results demonstrate that I-AdaIN, while maintaining movement rationality, significantly surpasses existing mainstream methods in style consistency, training efficiency, and anti-interference ability, achieving the optimal balance between dance movement style transfer effectiveness and practicality. All performance differences passed the significance test ($P < 0.01$), indicating statistical significance.

5 Discussion

The research proposes a dance action style transfer model based on an improved Ada IN algorithm, which achieves high fidelity transfer of fine-grained action styles. By introducing the spatiotemporal transformer structure and the joint limb global three-level layered style fusion mechanism, the accuracy and style consistency of action feature extraction are significantly improved. This method achieved a style detail retention rate of 0.89 and a JME of only 4.3mm, both of which were superior to those of the control group, demonstrating excellent detail preservation and structural integrity. Compared with the 'motion puzzle' method proposed by Koo et al. (2022) although it has the ability to control the style of body parts, it is still relatively coarse-grained in multi-level style modelling and overall style fusion. This study achieves a better balance between action style consistency and content fidelity through layered normalisation and transformer collaborative enhancement strategy. In addition, regarding the dynamic stability during style transfer, this study maintained a TSC score of 88.7 in the long sequence transfer task, far exceeding that of the sequential CNN, indicating its stronger modelling ability for dance rhythm and motion coherence. Compared to the style modulation animation synthesis system proposed by Mason et al. (2022) this study not only has higher efficiency in real-time (with a single-frame delay of only 8 ms), but also has higher controllability and physical rationality for the transfer effect of real dance actions, making it suitable for high dynamic rhythm Latin dance style scenes. In addition, in terms of cross-dataset generalisation ability, the research method still maintains a SSIM of 0.87 and a DAE of 0.12 on the CMU MoCap dataset without fine-tuning, demonstrating good transfer robustness. Especially when dealing with Latin dance actions with complex hip swing features, its multi-layer style modelling is superior to pose-based decoupling strategies such as PoseStyle, avoiding motion distortion caused by insufficient style feature expressions. Compared with the speech gesture style diffusion network proposed by Ao et al. (2023) this study enhances the structural rationality of style transfer actions through physical feasibility constraints and inverse kinematics correction modules, overcoming the 'drift' problem in high degree of freedom skeleton actions.

While research has significantly improved the fidelity, temporal consistency, and style control of dance style transfer, a bottleneck remains in the integration of dance styles with significant differences. This can be addressed by introducing cross-domain contrastive learning and style self-attention mechanisms to enhance the discriminative and expressive power of style features. Since the model relies on high-precision skeleton data, pose estimation errors in real-world scenarios may lead to fluctuations in results; further improvements in stability can be made by incorporating multi-view videos or robust filtering algorithms. Simultaneously, the model can be further optimised through lightweight structures and distillation generation modules to achieve efficient deployment on mobile devices and in AR/VR environments. Furthermore, the application of dance style transfer may raise ethical issues such as copyright and identity verification of dancers' performance styles. Future research should establish authorisation and traceability mechanisms during the data collection and model generation stages to ensure that technological innovation and artistic creation develop within a legal and controllable framework.

6 Conclusions

To improve the rationality of dance action style transfer methods on spatiotemporal feature coupling and physics, this study proposed a Latin dance action style transfer method based on improved Ada IN multi-feature fusion. The joint-limb-global three-level normalisation architecture was constructed to achieve fine-grained style decoupling, combined with inverse kinematics correction module to ensure the physical feasibility of actions, ultimately achieving high fidelity cross-dance style transfer. The experimental results showed that the improved method achieved SSIM of 0.94 in style transfer fidelity testing, which was 21%–27% higher than that of the original Ada IN (0.77) and

CycleGAN (0.74). The JME was reduced to 4.3 mm, with a decrease of more than 50%. In the content retention ability test, the MSE was 0.008, which was 50%–60% lower than that of the pure transformer (0.016) and VAE (0.020). The DTW distance (0.73) was 1.5 times better than that of the control group. SSIM and DAE maintained excellent performance of 0.87 and 0.12, respectively, in cross-dataset testing. In terms of system performance, the real-time processing speed reached 120 FPS, the single-frame processing time was controlled within 8 ms, and the TSC score was still 88.7 under long sequence migration (100,000 frames). The proposed method has significant advantages in style fidelity, content retention, real-time performance, and cross-domain adaptability, providing an effective technical solution for dance action style transfer.

Declarations

The author declares no competing interests.

References

Alexanderson, S., Nagy, R. and Beskow, J. (2023) 'Listen, denoise, action! audio-driven motion synthesis with diffusion models', *ACM Transactions on Graphics*, July, Vol. 42, No. 4, pp.1–20, DOI: 10.1145/3592458.

Ao, T., Zhang, Z. and Liu, L. (2023) 'Gesturediffuclip: gesture diffusion model with clip latents', *ACM Trans. Graph*, July, Vol. 42, No. 4, pp.1–18, DOI: 10.1145/3592097.

Chen, F., Wang, Y., Xu, S., Wang, F. and Sun, F. (2023) 'Style transfer network for complex multi-stroke text', *Multimedia Systems*, January, Vol. 29, No. 3, pp.1291–1300, DOI: 10.1007/s00530-023-01047-4.

Chen, J., Li, S. and Liu, D. (2022) 'Indoor camera pose estimation via style-transfer 3D models', *Comput.-Aided Civ. Infrastruct. Eng.*, June, Vol. 37, No. 3, pp.335–353, DOI: 10.1111/mice.12714.

Chen, Y., Yuan, Q., Li, Z., Xu, C. and Zhang, J. (2025) 'UPST-NeRF: Universal photorealistic style transfer of neural radiance fields for 3D scene', *IEEE Trans. Vis. Comput. Graph.*, March, Vol. 31, No. 4, pp.2045–2057, DOI: 10.1109/TVCG.2024.3378692.

Chung, C.Y. and Huang, S.H. (2023) 'Interactively transforming Chinese ink paintings into realistic images using a border enhance generative adversarial network', *Multimedia Tools and Applications*, August, Vol. 82, No. 8, pp.11663–11696, DOI: 10.1007/s11042-022-13684-4.

Garg, M., Ubhi, J.S. and Aggarwal, A.K. (2023) 'Neural style transfer for image steganography and destylization with supervised image to image translation', *Multimedia Tools and Applications*, August, Vol. 82, No. 4, pp.6271–6288, DOI: 10.1007/s11042-022-13596-3.

Hu, L., Zhang, Z., Ye, Y., Xu, Y. and Xia, S. (2024) 'Diffusion-based human motion style transfer with semantic guidance', in *Comput. Graph. Forum*, October, Vol. 43, No. 8, pp.e15169–e15170, DOI: 10.1111/cgf.15169.

Ji, Z. and Tian, Y. (2024) 'IoT based dance movement recognition model based on deep learning framework', *Scalable Comput.: Pract. Exp.*, February, Vol. 25, No. 2, pp.1091-1106, DOI: 10.12694/scpe.v25i2.2651.

Jiang, H. and Yan, Y. (2024) 'Sensor based dance coherent action generation model using deep learning framework', *Scalable Comput.: Pract. Exp.*, February, Vol. 25, No. 2, pp.1073–1090, DOI: 10.12694/scpe.v25i2.2648.

Khare, O., Mane, S., Kulkarni, H. and Barve, N. (2024) 'LeafNST: An improved data augmentation method for classification of plant disease using object-based neural style transfer', *Discov. Artif. Intell.*, July, Vol. 4, No. 1, pp.50–51, DOI: 10.1007/s44163-024-00150-3.

Khemakhem, F. and Ltifi, H. (2023) 'Neural style transfer generative adversarial network (NST-GAN) for facial expression recognition', *Int. J. Multimed. Inf. Retr.*, August, Vol. 12, No. 2, pp.26–27, DOI: 10.1007/s13735-023-00285-6.

Koo, J.D., Park, S. and Lee, S.H. (2022) 'Motion puzzle: arbitrary motion style transfer by body part', *ACM Trans. Graph*, June, Vol. 41, No. 3, pp.1–16, DOI: 10.1145/3516429.

Li, J., Wu, S., Zhang, X. and Luo, T.J. (2023) 'Cross-subject aesthetic preference recognition of Chinese dance posture using EEG', *Cognitive Neurodynamics*, June, Vol. 17, No. 2, pp.311–329, DOI: 10.1007/s11571-022-09821-2.

Liu, D.S. and Tu, N. (2021) 'Video cloning for paintings via artistic style transfer', *Signal Image Video Process*, July, Vol. 15, No. 1, pp.111–119, DOI: 10.1007/s11760-020-01730-3.

Mason, I., Starke, S. and Komura, T. (2022) 'Real-time style modelling of human locomotion via feature-wise transformations and local motion phases', *Proc. ACM Comput. Graph. Interact. Tech.*, May, Vol. 5, No. 1, pp.1–18, DOI: 10.1145/3522618.

Mukherjee, D., Saha, P., Kaplun, D. and Sinitca, A. (2022) 'Brain tumour image generation using an aggregation of GAN models with style transfer', *Sci. Rep.*, June, Vol. 12, No. 1, pp.9141–9142, DOI: 10.1038/s41598-022-12646-y.

Shih, C., Chen, Y. and Lee, T. (2021) 'Map art style transfer with multi-stage framework', *Multimed. Tools Appl.*, September, Vol. 80, No. 3, pp.4279–4293, DOI: 10.1007/s11042-020-09788-4.

Song, W., Jin, X. and Li, S. (2023) 'Finestyle: semantic-aware fine-grained motion style transfer with dual interactive-flow fusion', *IEEE Trans. Vis. Comput. Graph*, November, Vol. 29, No. 11, pp.4361–4371, DOI: 10.1109/TVCG.2023.3320216.

Tsuchida, S. (2024) 'Dance information processing: computational approaches for assisting dance composition', *New Generation Computing*, August, Vol. 42, No. 5, pp.1049–1064, DOI: 10.1007/s00354-024-00273-2.

Wiset, S. and Champadaeng, S. (2024) 'Lam in Ubon style: the process of transferring learning to inherit the performing arts', *Int. J. Educ. Lit. Stud.*, February, Vol. 12, No. 1, pp.121–125, DOI: 10.7575/aiac.ijels.v.12n.1p.121.

Yin, W., Yin, H., Baraka, K., Kragic, D. and Björkman, M. (2023) 'Multimodal dance style transfer', *Mach. Vis. Appl.*, Vol. 34, No. 4, pp.48–49, May 2023, DOI: 10.1007/s00138-023-01399-x.

Yin, W., Yu, Y. and Yin, H. (2024) 'Scalable motion style transfer with constrained diffusion generation', in *Proc. AAAI Conf. Artif. Intell.*, March, Vol. 38, No. 9, pp.10234-10242, DOI: 10.1609/aaai.v38i9.28889.

Yuk, J.M., Kim, Y.G. and Chun, J.Y. (2023) 'Study on performing arts in virtual environments using artificial intelligence and augmented reality', *Journal of Digital Contents Society*, November,Vol. 24, No. 12, pp.2981–2991, DOI: 10.9728/dcs.2023.24.12.2981.

Zhang, C., Xu, X. and Wang, L. (2024) 'S2wat: image style transfer via hierarchical vision transformer using strips window attention', in *Proceedings of the AAAI Conference on Artificial Intelligence*, March, Vol. 38, No. 7, pp.7024–7032, DOI: 10.1609/aaai.v38i7.28529.

Zhang, C., Xu, Y. and Mo, H. (2025) 'Semantic adjusted style transfer network with multi-attention mechanisms', *Comput. Eng. Appl.*, July, Vol. 61, No. 8, pp.204–214, 2025, DOI: 10.3778/j.issn.1002-8331.2311-0047.

Zhang, X., Yang, S., Xu, Y. and Zhang, W. (2023) 'Mining and applying composition knowledge of dance moves for style-concentrated dance generation', in *Proc. AAAI Conf. Artif. Intell.*, June, Vol. 37, No. 4, pp.5411–5419, DOI: 10.1609/aaai.v37i4.25673.

Zhang, Z., Zhang, Q. and Xing, W. (2024) 'Artbank: artistic style transfer with pre-trained diffusion model and implicit style prompt bank', in *Proceedings of the AAAI Conference on Artificial Intelligence*, March, Vol. 38, No. 7, pp.7396–7404, DOI: 10.1609/aaai.v38i7.28570.

Zhao, Y. and Yang, H. (2023) 'Implementation of computer aided dance teaching integrating human model reconstruction technology', *Computer-Aided Design and Applications*, March, Vol. 21, No. S10, pp.196–210, DOI: 10.14733/cadaps.2024.S10.196-210.

Zhou, Q., Li, M. and Zeng, Q. (2023) 'Let's all dance: enhancing amateur dance motions', *Comput. Vis. Media*, March, Vol. 9, No. 3, pp.531–550, DOI: 10.1007/s41095-022-0292-6.