

## Research Article

# A Cyclic Consistency Motion Style Transfer Method Combined with Kinematic Constraints

Huajun Wang,<sup>1,2</sup> Dandan Du,<sup>1</sup> Junhuai Li <sup>1,2</sup> Wenchao Ji,<sup>1</sup> and Lei Yu<sup>1,2</sup>

<sup>1</sup>School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China

<sup>2</sup>Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an 710048, China

Correspondence should be addressed to Junhuai Li; [lijunhuai@xaut.edu.cn](mailto:lijunhuai@xaut.edu.cn)

Received 18 February 2021; Accepted 24 May 2021; Published 30 June 2021

Academic Editor: Bin Gao

Copyright © 2021 Huajun Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motion capture technology plays an important role in the production field of film and television, animation, etc. In order to reduce the cost of data acquisition and improve the reuse rate of motion capture data and the effect of movement style migration, the synthesis technology of motion capture data in human movement has become a research hotspot in this field. In this paper, kinematic constraints (KC) and cyclic consistency (CC) network are employed to study the methods of kinematic style migration. Firstly, cycle-consistent adversarial network (CCycleGAN) is constructed, and the motion style migration network based on convolutional self-encoder is used as a generator to establish the cyclic consistent constraint between the generated motion and the content motion, so as to improve the action consistency between the generated motion and the content motion and eliminate the lag phenomenon of the generated motion. Then, kinematic constraints are introduced to normalize the movement generation, so as to solve the problems such as jitter and sliding step in the movement style migration results. Experimental results show that the generated motion of the cyclic consistent style transfer method with kinematic constraints is more similar to the style of style motion, which improves the effect of motion style transfer.

## 1. Introduction

Motion capture technology is based on the principles of computer graphics, recording the human body motion process through motion capture devices [1]. When the motion capture system is performing motion capture, it can track the motion trajectory of the moving object in the three-dimensional space and obtain the motion information of the moving object in the three-dimensional space through calculation processing. It has high precision, high quality, and complete motion information when representing human movements. The combination of motion capture data and computer animation technology can realistically restore actions. In recent years, it has been widely used in movies, games, medical treatment, sports, and other fields [2, 3]. In the field of film production, the use of motion capture technology has been particularly successful. Many animated films that use motion capture technology have achieved good box office results. In “Alita: Battle Angel,” the use of motion capture technology to capture actors’ actions and expressions is

processed by computer animation technology, making it difficult for viewers to distinguish the boundary between reality and animation. In the field of game production, the application of motion capture technology makes the characters in the game more realistic. High-precision motion capture data ensures the fluency of fighting actions and brings better game experience to players [4]. In the medical field, Noitom’s “Dr. Joint” [5] uses motion capture technology to address postoperative rehabilitation problems of knee patients and helps rehabilitation training by recording the patient’s activity and gait data. In addition, in motion training, the motion capture system can capture the detailed sports situation of the athletes, so as to better analyze the problems of the athletes, and make corresponding adjustments to achieve better training goals.

With the widespread application of motion capture technology in film, animation, and other production fields [3], research institutions such as Carnegie Mellon University, the University of Edinburgh, and the University of Bonn have established huge human motion capture databases.

Due to differences in collection objects and collection sites, it is necessary to recollect motion capture data for the same type of action, resulting in low reusability of motion capture data and increasing the cost in practical applications. Data-driven motion synthesis is a key technology to realize the reuse of human motion data. Through the study of human motion synthesis methods such as motion style transfer, motion retargeting, and motion blending, based on the existing motion capture data, the motion data that meets the needs of users is synthesized [6]. The collection work that originally required repeated collection actions or even replacement of collection objects can now be reduced by human motion synthesis technology, saving a lot of manpower and material resources, and improving the production efficiency of movies, animations, etc. The editing and synthesis methods of motion capture data have high research and practical application value.

At present, deep learning provides great convenience for motion style transfer, without the need for complex data preprocessing. However, there are still two problems with the motion style transfer method based on deep learning: first, because the motion capture data is time series data, the pooling process of neural network reduces the temporal correlation of motion data when extracting motion features, resulting in the difference of motion of generated motion and content motion at the same time, and the phenomenon of generated motion lags relative to content motion; second, the reconstruction of motion features results in the missing of some motion data frames, which leads to some problems such as jitter and sliding step in the generated movement after the motion style migration. In this paper, by determining the training target of kinematic constraint loss function and combining kinematic constraint with cyclic consistent confrontation generation network, the problems of animation jitter and sliding step in the process of style transfer are solved and make the style of the generated motion and the style motion closer.

The paper contribution: by combining the cyclic consistent style transfer method with kinematic constraints (KC), the motion style transfer network based on the convolutional autoencoder is used as the generator, and the cyclic consistent generation adversarial network (CCycleGAN) is constructed to establish cyclical consistency constraints between generated motion and content motion to further improve the consistency of generated motion and content motion; introduce kinematic constraints, standardize generated motion, solve problems such as jitter and sliding in the result of style transfer, and improve motion style transfer effect.

## 2. Related Research

With the development and application of motion capture technology, human motion synthesis technology based on motion capture data has attracted more and more attention from researchers at home and abroad and has made considerable progress.

*2.1. Motion Blending.* Early motion data is mainly composed of high-level motion parameters such as joint angles and joint coordinates. Therefore, technologies in the field of image and signal processing are used in the design, modification, and adaptation of motion data. Human motion data is treated as a time series signal for editing or fusion. Troje [7] proposed a motion synthesis framework that encodes motion patterns and uses linear methods for motion analysis and motion synthesis. Shapiro et al. [8] proposed an interactive motion data editing method, which uses independent components to analyze the motion style in the motion data and reedit the motion data to change the motion style. Wang and Bodenheimer [9] used the linear mixing method to determine the transformation point on the motion sequence by calculating the optimal weight of the basic cost metric.

Since it is difficult to directly synthesize more complex or obviously different motion styles with the method of signal processing on motion data, in order to solve the problem of poor synthesis of complex motion, some scholars establish kinematic constraints during motion generation to achieve smooth processing of generated motion [10]. With the improvement of animation effect requirements, in order to deal with more complex motion data, nonlinear processing methods [11] are applied to motion capture data with complex structures.

*2.2. Methods Based on Statistics and Learning.* Some scholars use statistics and learning methods to analyze the motion data, extract representative motion features and motion patterns in the motion data, and change the motion mode by adding constraints to generate new motions while retaining the existing motion characteristics.

Matthew and Hertzmann [12] learn the motion pattern of each motion style from a set of motion data sequences containing multiple motion styles. Each motion sequence can have a different choreography, and each choreography element has a different style. Through learning it can identify the general arrangement elements in the sequence and use interpolation to synthesize new motion data according to the action choreography elements. Grochow et al. [13] proposed an inverse kinematic system based on a human pose learning model. Given a set of kinematic constraints, the poses that are most likely to meet these constraints can be generated. The system uses different motion data for learning, generates the probability distribution of the motion sequence pose, determines the probability of a motion pose in the motion pose space through the objective function, and matches the pose to generate a new motion.

*2.3. Motion Graph.* In 2002, Kovar et al. [14] first proposed the concept of "motion graph," through the relationship between different motion data constructed, search for the optimal path in the constructed motion graph, and synthesize a new motion sequence. Arikian and Forsyth [15] proposed a framework for synthesizing motion by editing motion capture data. They regard motion synthesis as a combination problem and combine them by randomly searching the hierarchical structure of motion graphs. Since motion graph can only combine and edit motion capture data to

meet user needs, Min and Chai [16] enrich motion capture datasets by mixing the same type of motion data or combining sketches. The construction of a motion graph requires more on the quantity and type of motion capture data in order to be able to express the changes of the entire motion, and the new motion generated finally depends too much on the existing motion dataset. Parametric motion synthesis is to add the human body's footing, speed, acceleration, and other parameters to the synthesis model, control the synthesis process, and improve the problems of animation jitter and foot sliding [17].

**2.4. The Deep Learning Approach.** At present, applying deep learning to human motion capture data has become the main method of motion style transfer. Deep learning is used to synthesize new data, and the framework based on deep learning automatically learns features from the dataset. Taylor et al. [18] applied restricted Boltzmann machines to synthesize animation. On this basis, Mittelman et al. [19] proposed a structured constrained Boltzmann machine to improve the animation reconstruction. Subsequently, Fragkiadaki et al. [20] used an autoencoder (AE) recursive decoder network, which is a recurrent neural network that combines deep learning with time dynamics and produces smooth interpolated motion while reducing slipping. To further improve the animation effect, Du et al. [21] used multisource large-scale motion datasets to construct a hierarchical recurrent neural network and synthesized smooth and natural motion animation. In the motion editing method proposed by Holden et al. [22], a single-layer convolutional autoencoder is used for feature extraction, which also shows a better ability to express motion data, which promotes the autoencoder (AE) using in motion synthesis. In motion style transfer, Zan [23] establishes a self-encoding network structure with three convolutional layers and establishes style constraints in the feature space to realize motion style transfer. A novel data-driven framework is present for motion style transfer [24], which supports style extraction from videos and learns from an unpaired collection of motions with style labels. In this paper, Yu et al. propose that style translation is an effective way [25] to transform adult motion capture data to the style of child motion. Our method is based on CycleGAN.

The deep learning method based on autoencoder (AE) improves the effect of motion data synthesis or motion style transfer. However, encoding the motion data will cause a certain amount of data loss, which leads to jitter and slipping in the result of motion style transfer. In this paper, the method of movement style transfer is studied by combining kinematic constraints to improve the reuse rate of motion capture data.

### 3. The Whole Process of Cycle-Consistent Motion Style Transfer Combined with Kinematic Constraints

Style motion and content motion have more similar motion features, so the generated motion can maintain a high level of consistency with the content motion. However, the collection of style motion is difficult and the types of actions are

relatively few. There are relatively few content motion and style motion with similar motion content. When performing motion style transfer, using content motion with similar actions for motion style transfer can improve the transfer effect; when the content of the content motion and the style motion is quite different, the generated motion can remain similar to the content motion at the same time. However, the generated movement has obvious motion lag and causes the movement direction to deviate. Therefore, it is necessary to improve the similarity between the generated motion and the content motion while maintaining a high style similarity between the generated motion and the style motion. The transfer of motion style mainly involves problems such as difficulty in extracting motion features, poor reconstruction of motion effects, and establishment of motion style constraints. The overall process of motion style transfer of motion feature extraction and motion reconstruction network is shown in Figure 1.

It can be seen from Figure 1 that in order to transfer a specific style motion to content motion, it is first necessary to extract the motion features from the input motion data and then reconstruct the motion from the motion features to make the reconstructed motion data consistent with the input motion data. In order to realize the transfer of motion style, the style of the reconstruction motion to establish reasonable constraint, guarantee the reconstruction motion in reserves the content motion, content motion has the style of motion style and outputs the generated motion of the motion style transfer. Therefore, the motion style transfer network has the same structure as the motion feature extraction and motion reconstruction network, and the parameters are shared. The motion style constraints are established in the hidden layer feature space of the network, the motion characteristics are adjusted, and the motion style transfer is realized through motion reconstruction.

Aiming at the data loss caused by the use of autoencoder to encode the motion data in the process of motion style transfer, a cyclic consistent (CC) style transfer method combined with kinematic constraints (KC) proposed in this paper mainly includes two steps: (1) construction a cyclic consistent generated adversarial network; (2) combined with kinematic constraints to establish a cyclic consistent style transfer model.

## 4. The Construction of a Cyclic Consistent Generated Adversarial Network

**4.1. Theoretical Basis.** In the field of image processing, it is possible to use the cyclic consistent generated adversarial network to convert two image sample domains (nonpaired image domains) with large differences in style and improve the effect of nonpaired image style transfer [26]. Figure 2 shows the cycle-consistent generated adversarial network model. The cyclic consistent generated adversarial network is used for image-to-image mapping learning, and the learning method uses unpaired images for style transfer.

First, there are two unpaired image sample spaces  $X$  and  $Y$  with different contents. The goal of generating an adversarial network is to learn the mapping from  $X$  to  $Y$ . This

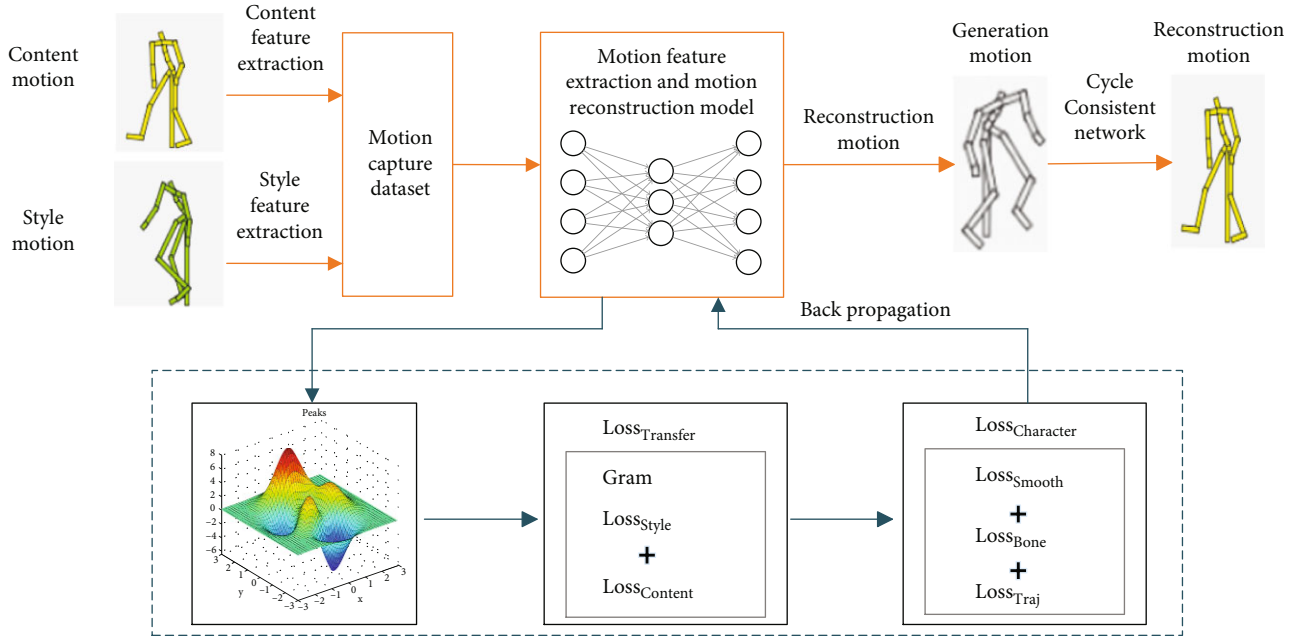


FIGURE 1: Motion style transfer process of motion capture data.

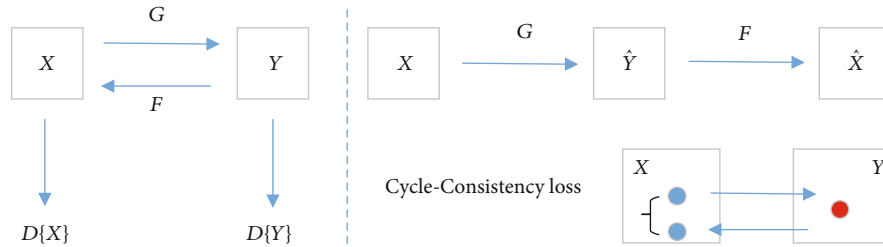


FIGURE 2: Cycle-consistent generative adversarial networks.

mapping is  $G$ , which corresponds to the generator in the generation adversarial network.

$$\hat{Y} = G(X, Y). \quad (1)$$

Among them, the generator  $G$  can convert the picture in the sample space  $X$  into a fake picture  $\hat{Y}$  similar to the image sample space  $Y$ , and it is hoped that the style of the generated image  $\hat{Y}$  and  $Y$  is as similar as possible.

For the generated picture  $\hat{Y}$ , the discriminator  $D\{Y\}$  is used to determine whether it is a real  $Y$  picture, thereby forming a generated adversarial network.

$$\text{Loss}_{\text{GAN}}(G, D, A, B) = \log D(Y) + \log (1 - D(G(X))). \quad (2)$$

Using only this one loss will cause the mapping  $G$  to map all the images in the sample space  $X$  to the same image in the  $Y$  space, invalidating the loss. Therefore, by introducing the mapping  $F$ ,  $\hat{Y}$  can be transformed into a picture  $\hat{X}$  similar to the sample space  $X$ .

$$\hat{X} = F(\hat{Y}, X). \quad (3)$$

And then establish the connection between  $\hat{X}$  and  $X$ , forming a circular consistency constraint.

$$\text{Loss}_{\text{Cyc}}(G, F, X, Y) = \|F(G(X)) - X\|_2^2. \quad (4)$$

The cyclic consistency constraint is applied to the image style transfer, and the information of the content image during the transfer can be retained as much as possible, so that the generated image after the transfer is more complete and natural. Applying it to motion style transfer can reconstruct the generated motion and establish the connection between the generated motion and the content motion through the cyclic consistency constraint, so that the generated motion retains more motion content and improves the consistency of the generated motion and the content motion and improves the effect of motion style transfer.

**4.2. Establishment of Cyclic Consistency Constraint.** In the motion style transfer, two sample spaces are defined: content motion  $C$  and style motion  $S$ . The cyclic consistency constraint (CC) is applied to the motion style transfer to establish a cyclic consistency style transfer model, as shown in Figure 3.

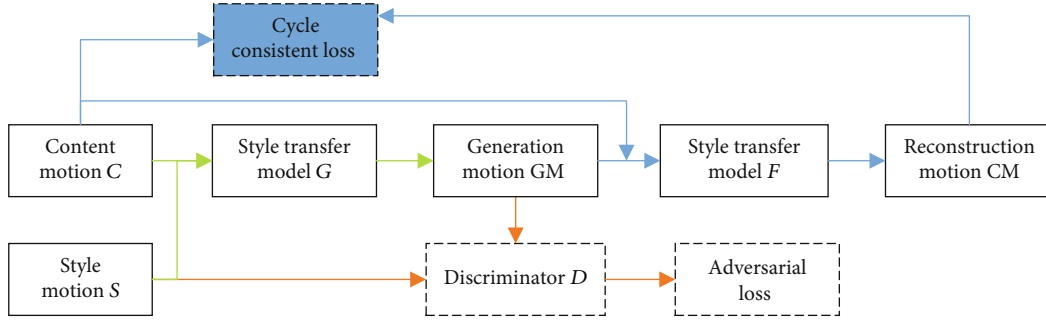


FIGURE 3: Cyclic-consistent style transfer model.

In the motion style transfer model, generators  $G$  and  $F$  are all motion style transfer models based on motion capture data. The content motion  $C$  and the style motion  $S$  are used as inputs of the generator  $G$  to perform motion style transfer to obtain the generated motion  $GM$ .

$$GM = G(C, S). \quad (5)$$

The discriminator  $D$  judges the style difference between the generated motion  $GM$  and the real style motion  $S$  through equation (6), so that the motion style of the generated motion is close to the motion style of the input  $S$ .

$$\text{Loss}_{\text{Style}} = \|\text{Gram}(H_s) - \text{Gram}(R(H_s))\|_2^2. \quad (6)$$

The motion feature  $H_s$  of the input style motion  $S$  and the motion feature  $H_{GM}$  of the generated motion  $GM$  are obtained through the discriminator  $D$  to obtain the confrontation loss:

$$\text{Loss}_{\text{GAN}} = \log D(H_s) + \log (1 - D(H_{GM})). \quad (7)$$

Generative adversarial networks generally measure the generation effect through the log loss function [27].

The use of the discriminator for training will emphasize the features of the motion style, making it difficult for the generator to retain the motion content and structure of the content motion, and it is necessary to add cyclic consistency constraints to encourage the content of the motion to be retained in the alignment process. The generated motion  $GM$  is transferred to the generation network  $F$ , the motion style of the content motion  $C$  is transferred to the generated motion  $GM$ , the motion style of the style motion  $S$  in the generated motion  $GM$  is removed, and the reconstructed motion  $CM$  of the generated motion  $GM$  is obtained.

$$CM = F(GM, C). \quad (8)$$

The paradigm  $L2$  is used to establish the consistent loss of the reconstructed motion  $CM$  and the content motion  $C$ , thereby effectively achieving cyclic consistency, so that the generated motion  $GM$  has more motion features of the content motion during reconstruction:

$$\text{Loss}_{\text{cyc}} = \|CM - C\|_2^2. \quad (9)$$

The cyclic consistency constraint makes the generated motion after the style transfer reconstitutes the original input content motion. By establishing the cyclic consistency constraint, the transferred motion style in the generated motion is removed to form a cycle. In this way, let the network learn the process of motion style transfer and then remove, so that the generated motion has more content motion features.

## 5. Cyclic Consistent Style Transfer Method Combined with Kinematic Constraints

The cyclic consistency constraint can establish the connection between the content motion and the generated motion and enhance the consistency of the generated motion and the content motion. However, due to the complex structure of the motion capture data, the inheritance relationship between the bone joint points makes the motion data highly correlated, and the data of the autoencoder is lossy, resulting in a gap between the motion generated after the encoding and decoding operation and the content motion. The resulting motion data frame is unreasonable, the action content is incomplete, and problems such as jitter and sliding footsteps occur.

At present, the method of solving motion jitter and foot slippage in motion synthesis is to add constraints to the motion synthesis results and standardize the motion data. By establishing kinematic constraints (KC), dynamic constraints and spatiotemporal constraints, and other methods [28], constraints are added to the generated motion to obtain complete and smooth and natural motion data. Lee and Shin [29] first used inverse kinematics to establish kinematic constraints for each frame of motion data and used multilevel spline curve interpolation to achieve smooth complete motion. Tak and Ko [30] added dynamic constraints on the basis of the previous kinematic constraints and transformed the spatiotemporal optimization problem into a constraint state estimation problem. Choi and Ko [31] based on inverse kinematics to calculate the joint angle change from the position of the extremity to realize the editing of motion data. Gleicher [32] realized motion synthesis based on spatiotemporal constraints but did not consider the kinematics and dynamic constraints of the generated motion and lacked the reality of motion. Zhang et al. [33] propose a motion retargeting method based on spatiotemporal constraints, which imposes spatiotemporal constraints on joint positions to avoid unreasonable motion. The kinematic constraints

established by Grochow et al. [13] use end effectors to clarify the position that the extremity needs to reach. Zhou et al. [34] construct a variety of kinematic constraints to edit motion data to realize motion retargeting.

Among kinematic constraints (KC), dynamic constraints, and spatiotemporal constraints, the motion synthesis method based on spatiotemporal constraints is computationally expensive and time-consuming. Dynamic constraints require fine-grained parameter control of the motion frame. Usually, dynamic parameters such as speed and acceleration are used to directly modify the motion features. It is necessary to rely on experience to achieve parameter control, and the generated results are uncertain. Kinematic constraints are further constraints on the consistency between the generated motion and the content motion and are targeted constraints. By determining the kinematic constraints to generate the motion, it provides a more reasonable generated motion for the cyclic consistent constraint and improves generation of the consistency of motion and content motion. Kinematic constraints have a large range of optional constraints and strong applicability. In this paper, three common constraints, such as smooth constraint, bone length constraint, and trajectory constraint, are selected for smoothing processing in character animation. Kinematic constraint loss function training objective is determined to combine kinematic constraint with cyclic consistent resistance generation network to solve the problems of jitter and sliding. The style transfer model is shown in Figure 4.

The kinematic constraints in the style transfer model shown in Figure 4 mainly include three aspects:

#### (1) Motion smoothing constraint

Villegas et al. [35] found that the data frames of continuous motion are highly dependent on the previous and subsequent data frames when performing motion retargeting, that is, the motion of each frame in the motion data is slightly changed compared with the motion of the previous frame, which can be generated by generating motion and content motion. The speed changes of the front and back data frames are used to the motion smoothly constrain and solve the problem of generating motion sliding. The smoothing constraint is defined as follows:

$$\text{Loss}_{\text{Smooth}}(H) = \sum_j \|v_j - v_{j-1}\|^2 - \|v_j' - v_{j-1}'\|^2. \quad (10)$$

Among them,  $v_j'$  is the motion speed of the joint point  $j$  in the three-dimensional space coordinate system in the content motion, and  $v_j$  is the motion speed of the joint point  $j$  in the three-dimensional space coordinate system that generates the motion.

#### (2) Bone length constraint

The motion capture data collected by the same motion capture device has the same bone hierarchy, but the length of the bones is not the same. The generated motion data needs to be consistent with the bone data of the content

motion. When Villegas et al. [35] use motion features to reconstruct motion data, it uses bone length constraints to ensure that the bones that generate motion will not be deformed and avoid the jitter of generated motion. This paper uses the three-dimensional space coordinates of the joint points as input data and imposes bone length constraints between adjacent joint points to maintain the stiffness of the body, so that the movement body that generates the motion will not cause movement dislocation due to deformation. The loss function of the bone length constraint is defined as follows:

$$\text{Loss}_{\text{Bone}}(H) = \sum_i \sum_b \|p_{b1}^i - p_{b2}^i - l_b\|^2. \quad (11)$$

Among them,  $i$  represents the number of motion frames,  $b$  represents the number of human bones, and  $p_{b1}^i$  and  $p_{b2}^i$  are the two end joint points that generate a segment of bone in motion.  $l_b$  is the length of bone  $b$ .

#### (3) Motion trajectory constraint

The motion style transfer hopes that the generated motion follows the motion trajectory of the content motion, so that the motion postures of the generated motion and the content motion are synchronized. Therefore, the motion needs to be precisely restricted to a certain trajectory. Holden et al. [22] edit and generate the motion data of the given trajectory through the high-level motion parameter of the given trajectory and generate the trajectory route of the motion through trajectory constraints. The motion trajectory constraint loss function is defined as follows:

$$\text{Loss}_{\text{Traj}}(H) = \|w - w'\|^2 + \|v_r - v_r'\|^2. \quad (12)$$

Among them,  $w$  is the axis angular velocity of the generated motion around the  $y$  axis,  $w'$  is the axis angular velocity of the content motion,  $v_r$  is the motion speed of the generated motion root joint point, and  $v_r'$  is the motion speed of the content motion root joint point.

Therefore, the loss of the three kinematic constraints combined with smoothing constraint, bone length constraint, and trajectory constraint is defined as follows:

$$\text{Loss}_{\text{KC}} = \arg \min_H \text{Loss}_{\text{Smooth}}(H) + \text{Loss}_{\text{Bone}}(H) + \text{Loss}_{\text{Traj}}(H). \quad (13)$$

Through training, kinematic constraint loss is minimized to obtain kinematic constraint motion features of the hidden layer, which can constrain the joint to the desired position while maintaining the stiffness of each bone. The kinematic constraints are adjusted to generate the motion backpropagation to the feature  $H$  of the hidden layer, until the hidden layer feature that can minimize the kinematic constraint value is obtained to realize the kinematic constraint. Therefore, the overall transfer loss of the circular consistent style transfer model combined with kinematic constraints is as follows:

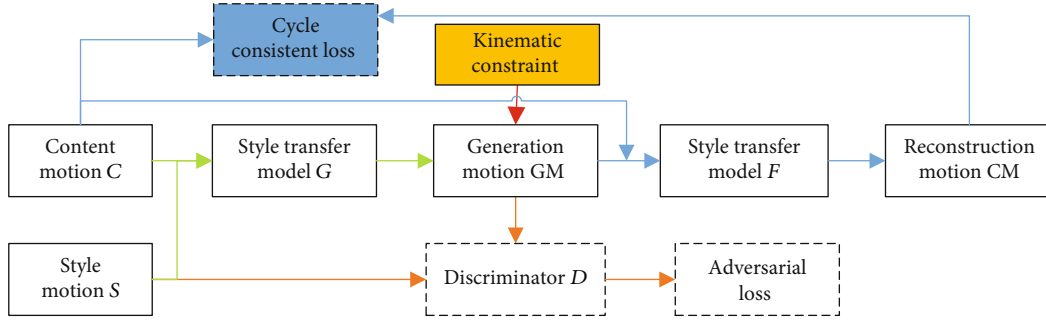


FIGURE 4: Cyclic-consistent style transfer model with kinematic constraints.

$$\text{Loss}_{\text{cycGAN}} = \text{Loss}_{\text{KC}} + \min_G \max_D \text{Loss}_{\text{GAN}} + \text{Loss}_{\text{cyc}}. \quad (14)$$

Among them,  $L_{\text{Kine}}$  is the kinematic constraint loss,  $\text{Loss}_{\text{GAN}}$  is the adversarial loss, and  $\text{Loss}_{\text{cyc}}$  is the cyclic consistency loss.

The specific network structure of the cyclic consistent style transfer model combined with kinematic constraints is shown in Table 1. The generated network takes the content motion and the style motion as the input of the generator to get the generated motion, and the discriminant network judges the style difference between the generated motion and the real style motion, so that the generated motion style is close to the input motion style. The two-generation networks have the same network structure as the motion style transfer network based on convolutional autoencoding, and the network parameters are shared. The discriminant network structure and motion feature extraction are the same as the coding network structure of the motion reconstruction model, sharing network parameters.

## 6. Experiment and Analysis

**6.1. Data Processing and Model Training.** In this paper, the CMU motion capture dataset [36] was used, with 2600 sets of BVH motion data of about 3 million frames as training data. The input motion capture data dimension is 73, consisting of 21 bone joints, plus the initial coordinate value of the bone root joint to form the data dimension. In order to process input data with large dimension, the number of hidden units in the neural network is generally more than twice of the data dimension, that is, the number of hidden units should be greater than 146, so the number of hidden units is set as 256. The collection frequency of the motion capture dataset in this paper is 120 frames per second. A human body motion will last for about 1 second to 2 seconds. In order to maintain a certain degree of integrity and continuity of the motion data of each batch, all motion data will be split once every two seconds and retain the action content of the second one after the previous motion, that is, the data is split by a 50% overlap window of 240 frames, and the motion data segment with less than 240 frames is filled with the last frame of the current data segment. Therefore, the size of the filter in the network is  $5 * 5 * 1$ , corresponding to about half a second of motion data, which is a reasonable sequence length for

most motion [11]. Due to the large training dataset, in order to improve the training speed, gradient descent uses Adam to update the parameters  $W_0$  and  $b_0$  and sets the learning rate  $\alpha = 0.01$  [37]. Use all datasets for 100 complete training, that is, epoch = 100, and use all the data to update the network parameters in each backpropagation, that is, batchsize = 1. The initialization of  $W_0 \in R^{m*d*w_0}$  selects a small random value, and the initialization of  $b_0 \in R^m$  is 0.

Although the Euler angle representation method of BVH motion capture files is intuitive and convenient in animation playback, the Euler angle rotation component performs poorly in characterizing the spatiotemporal characteristics of motion. Therefore, this paper first transforms the BVH Euler angle data into three-dimensional space coordinates of joint points.

At the same time, in order to make the trained network have better stability, the motion capture data is normalized when constructing the dataset used for autoencoder training [22]. All spatial coordinates in BVH dataset are normalized as follows:

$$X = \frac{X_{\text{input}} - X_{\text{mean}}}{X_{\text{std}}}, \quad (15)$$

where  $X_{\text{input}}$  is the input motion capture dataset,  $X_{\text{mean}}$  is the average value of the input motion data,  $X_{\text{std}}$  is the standard deviation of the input motion data, and  $X$  is the standard motion data after processing.

This paper selects three styles of old people, zombies, and orangutans and transfers them to ordinary people's walking and running, respectively, and analyzes the effect of the cyclic consistent style transfer method combined with kinematic constraints. That is, the three kinds of motions of old people, zombies, and orangutans are style motions, and ordinary people's walking is the content motion. The style motion and content motion are used as the input data of the network to transfer the motion style to obtain the generated motion.

**6.2. Experimental Analysis Process.** This paper is mainly from the following six aspects of the experiment.

- (1) Analysis on the results of motion style transfer

The migration results of ordinary people's walking and the three styles of motion are shown in Table 2. The three

TABLE 1: Cyclic-consistent style transfer networks with kinematic constraints.

	Layer	Shape	Param
Generated network	Input	(none, 256, 73)	
		Encoder network	
	DropoutLayer		Dropout = 0.25
	Conv1DLayer	(none, 73, 240)	
	Pool1DLayer	(none, 256, 240)	
		Decoder network	
	Depool1DLayer	(none, 256, 240)	
	DropoutLayer		Dropout = 0.25
	Conv1DLayer	(none, 256, 240)	
	Output	(none, 256, 73)	
Discriminant network		Adversarial network	
	DropoutLayer		Dropout = 0.25
	Conv1DLayer	(none, 73, 240)	
	Pool1DLayer	(none, 256, 240)	

styles of motion are transferred to ordinary people's walking movement through the cyclic consistent style transfer model combined with kinematic constraints. Compared with the style transfer results without constraints, the motion poses of the three style motion transfer results are closer to those of the content motion, and the style transfer effect is good.

It can be seen by performing the style transfer of the old people with or without constraints on the same posture that the difference in the effect of the constrained style transfer compared with the unconstrained style transfer is mainly reflected in the processing of complex motions, especially in the last column of the turning action; the constrained style transfer can maintain high consistency between generated motion and content motion. In contrast, unconstrained style transfer, although the motion posture is close to the content motion, is affected by the style motion, and there is room for improvement in the relative positions of joints and human body orientation; the obvious difference in the results of zombie style motion with and without constraint style transfer is the orientation of the generated motion and the content motion posture. In this group of forward and turn motions, the generative motion of unconstrained style transfer is relatively lagging. When the content motion turns, the generated motion is still going straight. Obviously, the style transfer effect with kinematic constraints is better for the posture constraints at the same time; the style transfer result of the orangutan style motion is with or without constraints, although the generative motion is with unconstrained style transfer. The posture and orientation are close to the content movement, but the relative positions of the feet are different compared to the content motion. In the constrained style transfer result, the relative positions of the feet of the generated motion are clearer and the footing point is more clear.

### (2) Analysis on the effect of cyclic consistent constraint and kinematic constraint

In order to verify the effect of cyclic consistent constraint and kinematic constraint, the motion style transfer result (unconstrained transfer result) based on convolutional auto-encoder is compared with the cyclic consistent style transfer result (constrained transfer result) combined with kinematic constraint.

The results of motion style transfer on the same level are mainly analyzed by observing the posture and footsteps of the human body. From the migration results of unconstrained migration and constrained migration in Table 3, it can be seen that the constrained generated motion and content motion in the style transfer of old people walking are more consistent in posture; the footsteps of generated motion in the style transfer of zombie walking. The horizontal contact is normal, which effectively solves the problem of ground penetration; the generated motion in the style transfer of the orangutan walking can move according to the position of the content motion.

### (3) The trajectory of motion style transfer results

In order to further illustrate the effect of motion style transfer, the motion trajectory diagram is visualized. Figure 5 is a trajectory diagram of ordinary people walking and three styles of motion.

The following moves the walking movement of ordinary people to the three styles of old people, zombie, and orangutan. The results of the migration trajectory are shown in Table 4.

Compared with unconstrained style migration, the trajectory diagram of constrained style migration is closer to the content motion and the trajectory is more complete. Compared to the content motion, generated with constraint style migration movement trajectory distance is shorter, and generated trajectory diagram is small; this is mainly due to the old style, zombie style speed slowly, leading to generated movement of the whole movement distance is short,



TABLE 2: Comparison of unconstrained and constrained transfer results between ordinary people walking and the three styles of motion.

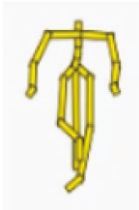
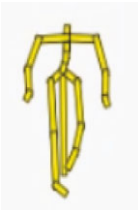






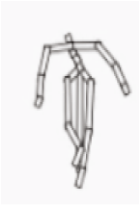

















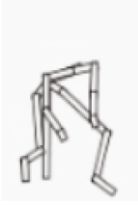






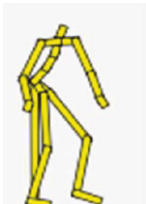








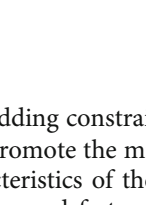
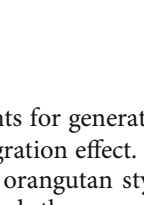
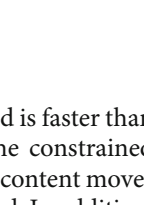
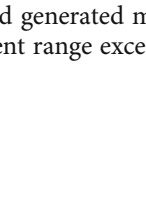
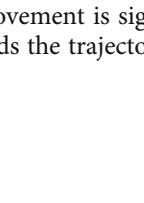
Motion type		Motion I	Motion II	Motion III	Motion IV
Content motion: ordinary people's walking					
	Unconstrained				
Walking style of the old people	Constrained				
	Unconstrained				
Walking style of zombies	Constrained				
	Unconstrained				
Walking style of orangutans	Constrained				

TABLE 3: Results of unconstrained and constrained movement style migration.

Behavior	Content motion	Unconstrained migration	Constrained
Old people—walking migration			
			
Zombies—walking migration			
			
			
			
Orangutans—walking migration			
			

The generated motion poses are similar, but the orientation is obviously different

Although the generated motion can maintain a posture similar to the content motion, the motion is lagging

Motion lag

The overall position of the generated motion at the same level is lower than that of the content motion, and the right foot penetrates the horizontal plane

The problem of large distance between footholds is prone to occur, leading to slipping

but from the whole, adding constraints for generated movement can effectively promote the migration effect.

Due to the characteristics of the orangutan style movement with large strides and fast speed, the movement distance of unconstrained generated movement is significantly larger, so the movement range exceeds the trajectory collec-

tion range. However, since the movement speed is faster than that of the old man and the zombie style, the constrained generated movement trajectory is closer to the content movement trajectory, and the migration effect is good. In addition, the constrained generated motion is significantly larger in distance between the feet and the stride length compared

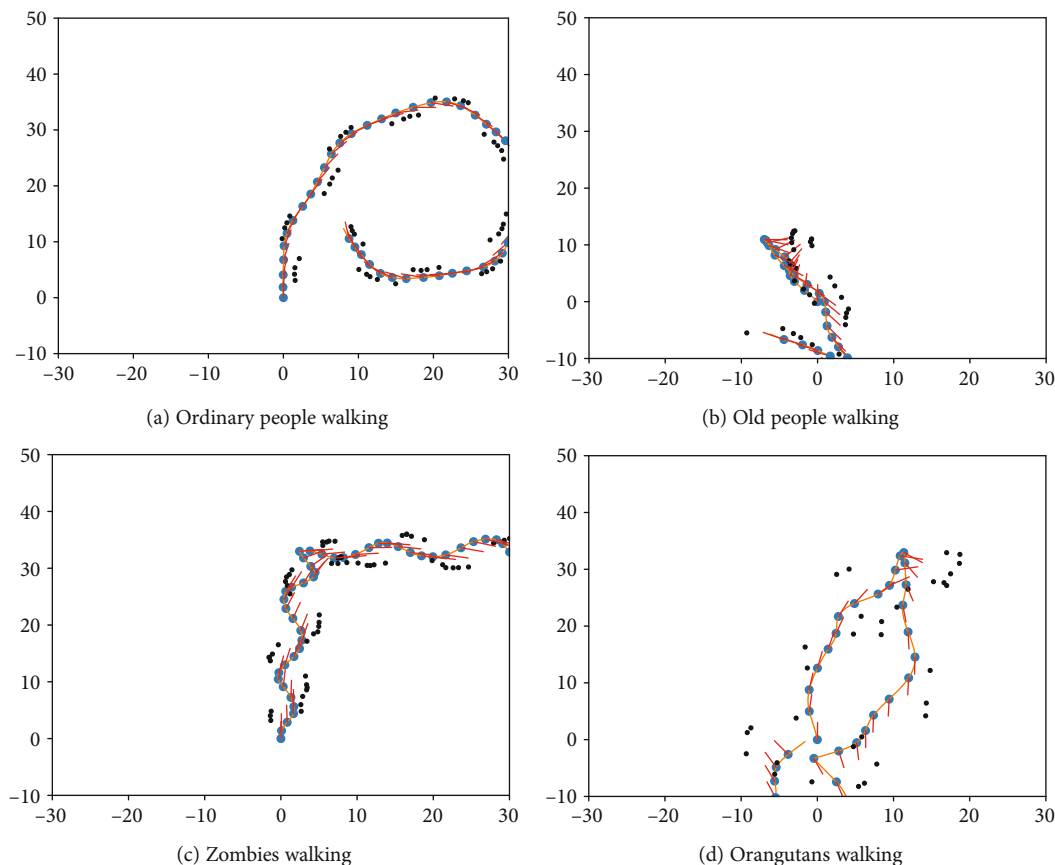


FIGURE 5: Content motion and style motion trajectories.

with the content motion, which is also in line with the characteristics of the large stride length of the orangutan style, indicating that the style feature retention effect is good.

According to the above analysis, unconstrained style transfer is prone to problems such as jitter and slippage caused by unclear foothold position, foot end penetration, and motion lag. The constrained style transfer can effectively solve the constraints on joint positions and motion trajectories, thereby further improving the effect of style transfer.

#### (4) Analysis of the style transfer results of unpaired motion data

In order to verify the style transfer effect of the unpaired input motion data, this paper takes the running motion that is different from the content motion of the three styles as the content motion. Due to the cyclic consistency constraint, the training is carried out by reducing the difference between the generated motion and the content motion. Therefore, the constrained generated motion should be closer to the posture of the content motion than the unconstrained generated motion. The following is to judge the consistency of the motion posture of the generated motion and the content motion based on the footsteps in the trajectory diagram and transfer the three styles of motion to running. The trajectory diagram of the input motion data is shown in Figure 6. It shows the running motion of ordinary people. The walking

motion of old people, the walking motion of zombies, and the walking motion of orangutans are the same as in Figure 5.

The following three styles of motions are transferred to ordinary people's running motions through a circular consistent style transfer model combined with kinematic constraints. The migration results are shown in Table 5.

The unconstrained style transfer results of the old people's style motions are poor. Compared with the content motion trajectory graph, the footsteps and the ground are in intensive contact, indicating that the motion content of the generated motion and the content motion is quite different. The generated motion trajectory of the constrained style transfer is closer to the trajectory diagram of the content motion, indicating that the cyclic consistency constraint still has a better migration effect on the style motion and the content motion that have different action content; zombie, orangutan style motion, and running motion are quite different, so there is a big difference between the unconstrained generated motion trajectory graph and the content motion trajectory graph, and it is impossible to determine the starting position of the movement and the similar trajectory paragraph. Constrained generated motion trajectory graphs still maintain a good migration effect, and the trajectories of generated motion and content motion are highly similar.

According to the results of the migration of the three styles of movement to the running movement, compared with the content movement, the generated movement of

TABLE 4: Results of moving movement migration.

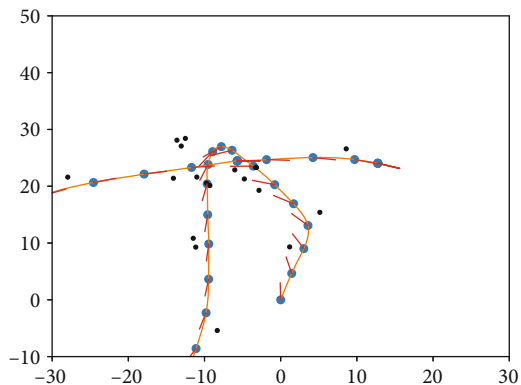
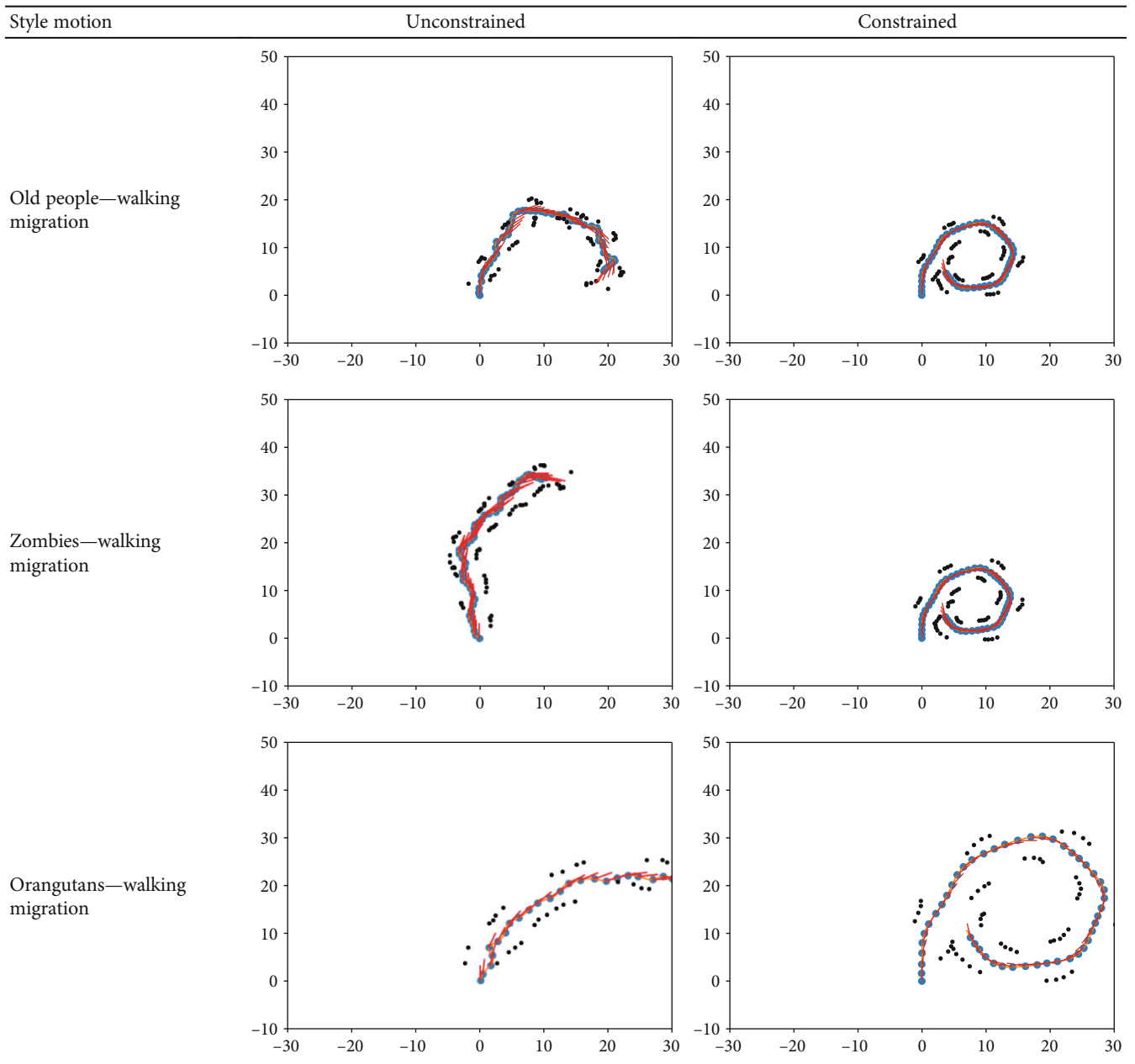


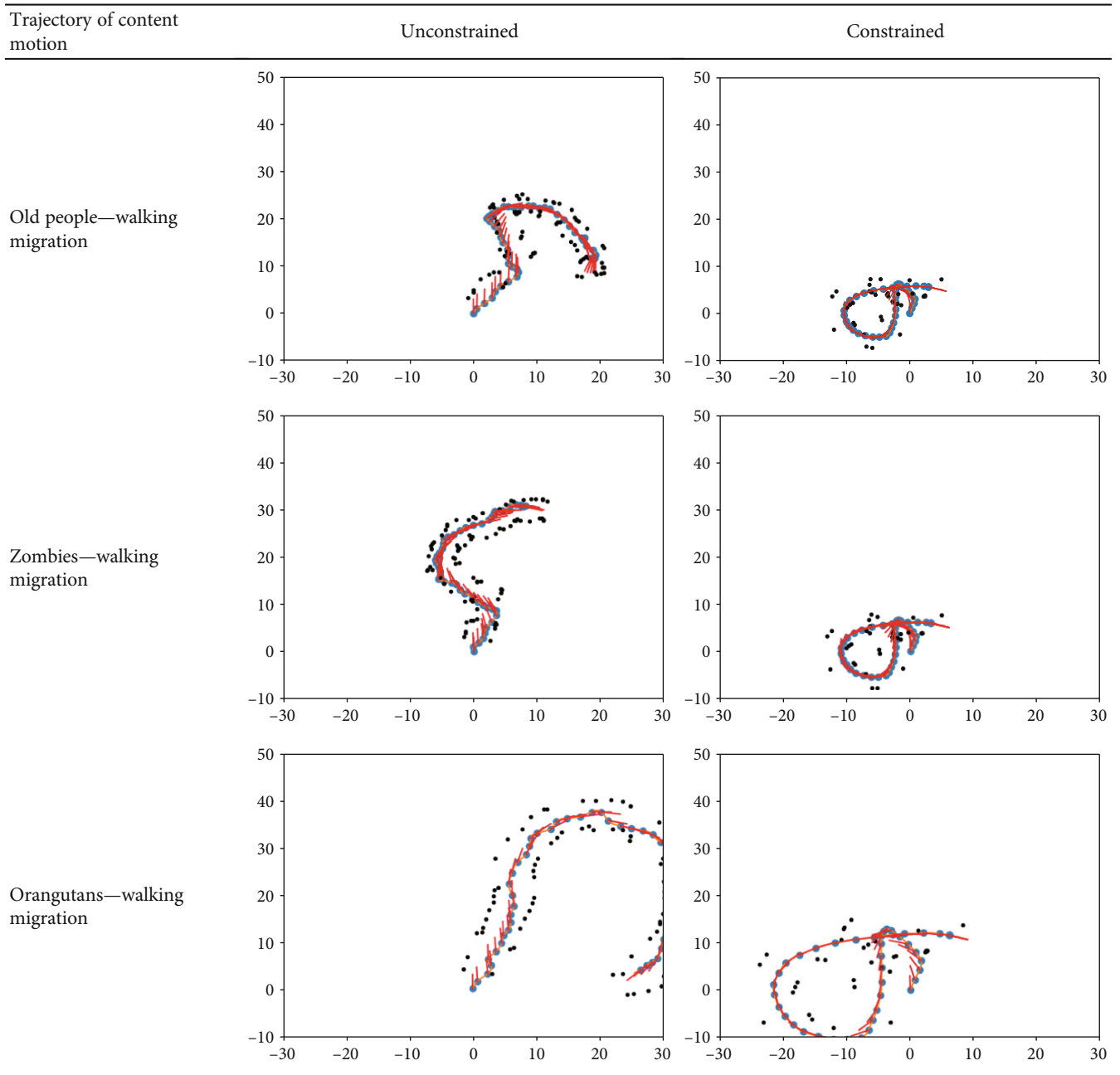
FIGURE 6: The ordinary human being running trajectory.

the unconstrained style transfer has lower similarity between the movement posture and the movement trajectory, and it is difficult to judge similar trajectory paragraphs. The cyclic consistent style transfer method combined with kinematic constraints has a good overall transfer effect. The generated motion retains style characteristics while ensuring a high degree of similarity of motion posture and motion trajectory.

(5) Motion style transfer training loss

Figures 7–9 show the changes in the loss values of the three styles of motion transfer to walking and running through the migration model. The solid line represents the

TABLE 5: Style migration results of unpaired motion data.



change of the migration loss value of the walking content motion, and the dotted line represents the change of the migration loss value of the running content motion.

In Figures 7–9, loss value with the increase of the number of iterations quickly converge, and the top 50 iteration training effect is relatively obvious; since the number of iterations is 100 times, loss value change is leveling off, and three style movement of migration loss value variation that can be seen; as the movement style complexity increases, the initial loss value is more and more big, and in the old man and the zombie migration

style, loss value is changing, due to differences in content and style motion increases, leading to loss of migration to the running value slightly higher than the migrated to loss value of the walking motion. Since the action content of orangutan style was close to that of running, there was no significant change in the loss value of moving to the two content motions.

(6) Evaluation of similarity of movement style transfer

Through equations (16) and (17), we calculated the style similarity of the generated movement from three styles of



FIGURE 7: Changes of old man-run style translation loss.

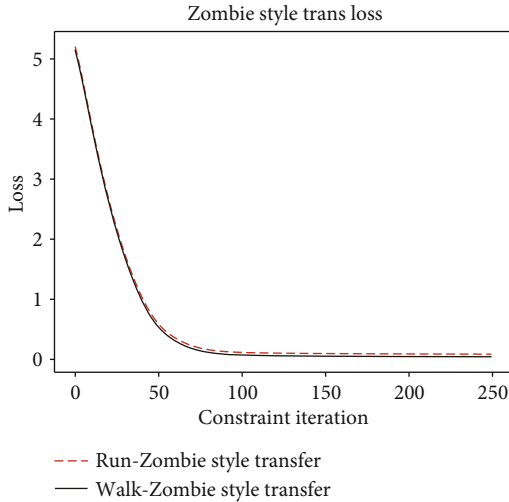


FIGURE 8: Changes of zombies-run style translation loss.

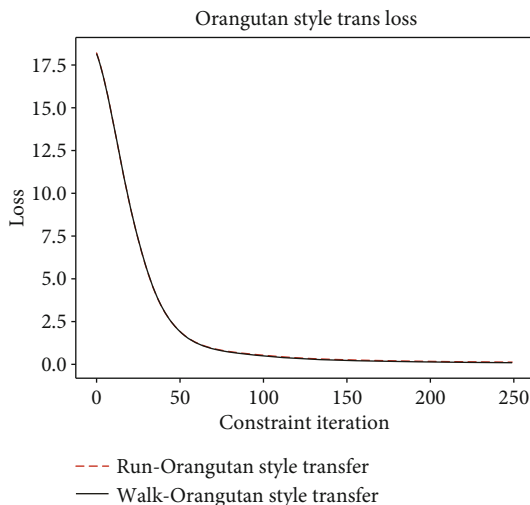


FIGURE 9: Changes of orangutan-run style translation loss.

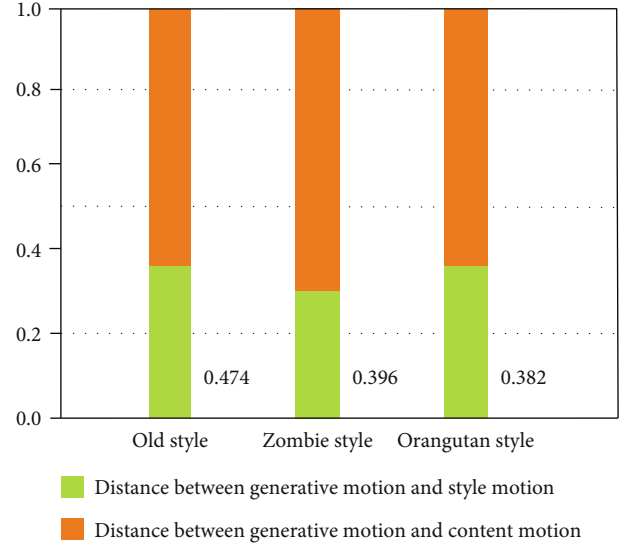


FIGURE 10: Style translation similarity.

motions to running.

$$D_S = \frac{D_{\text{style}}}{D_{\text{style}} + D_{\text{content}}}, \quad (16)$$

$$D_C = \frac{D_{\text{content}}}{D_{\text{style}} + D_{\text{content}}}, \quad (17)$$

where  $D_S$  is the style similarity between the generated movement and the style movement, and  $D_C$  is the style similarity between the generated movement and the content movement.

From the calculation results of style similarity in Figure 10, it can be seen that the style similarity value of the generated motion and the style motion of the circular consistent style transfer method combined with the kinematic constraint is lower than the style similarity value of the generated motion and the content motion, indicating the generated motion similar to the motion styles of the three styles of motions. In order to ensure the migration effect, the kinematic constraint and the cyclic consistency constraint are added, so that the generated motion and the content motion are more similar in the action content. In addition, the action content of running and the three styles of motion is quite different, while the action content of the walking and the three styles of motion is small, which leads to an increase in the style similarity value of the generated motion and the content motion.

In order to compare the effect of style transfer, this paper compares with Zan [23], Hu [38], Guo et al. [39], and Holden et al. [22]. Guo et al. [39] proposed a style transfer method combined with inverse kinematic constraints. First, the motion sequence is aligned through dynamic time warping, and then, the motion sequence is edited by establishing inverse kinematic constraints to realize the motion style transfer. Holden et al. stack a feedforward neural network on a single-layer convolutional autoencoder and edit the

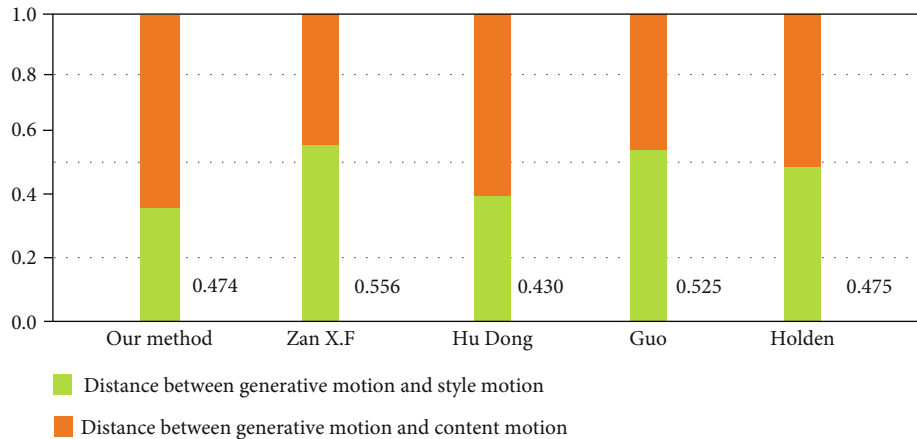


FIGURE 11: Style translation similarity comparison.

motion sequence through high-level parameters to generate the target motion sequence. Through the above five methods, the elderly style motions are transferred to running motions, and then, the style similarity is compared. From the calculation results of style similarity in Figure 11, it can be seen that the generated movement and content movement of Zan [23] and Guo et al. [39] are more similar, and the effect of style transfer is not ideal. The method of this paper and Hu [38] generates relatively small values of similarity between the movement and the style movement, and the effect of style transfer is better.

## 7. Conclusion

Aiming at the problem of generated motion postures lagging behind content motion at the same time and generated motion jitter sliders in the motion style migration method based on convolutional autoencoder, a cyclic consistent style migration method combining kinematic constraints is proposed. By constructing a cyclic consistent generation adversarial network, the motion style transfer network based on the convolutional autoencoder is used as a generator to establish a cyclic consistency constraint between the generated motion and the content motion, which improves the consistency of the generated motion and the content motion, and eliminates generated motion lagging. Kinematic constraints are introduced to standardize the generation of motion, which solves the problems of jitter and sliding in the results of motion style transfer and improves the effect of motion style transfer.

In the cyclic consistent style transfer model combined with kinematic constraints, physical factors are not considered, and physical constraints on generated motion are lacking. For example, when constraining the position of joint points, it did not consider whether the matching between footing point and motion speed is reasonable after the motion style transfer, the change of human muscles, gravity, and other factors. Reasonable physical constraints are also a way to improve the effect of generated motion, which is planned to be the content of subsequent research.

## Data Availability

The data used to support the findings of this study have been deposited in the CMU, Carnegie-Mellon Mocap Database repository (<http://mocap.cs.cmu.edu/>).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

H.J.W and J.H.L performed the conceptualization and methodology; D.D.D and W.C.J contributed to the software and validation; W.C.J helped in the original draft preparation; D.D.D and L.Y wrote, reviewed, and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

## Acknowledgments

This work was funded by the National Key R&D Program of China (No. 2017YFB1402103), Natural Science Foundation of China (No. 61971347), Scientific Research Program of Shaanxi Province (2016KTZDNY01-06 and 2018HJCG-05), Shaanxi Water Conservancy Technology Project (2020slkj-17), and Project of Xi'an Science and Technology Planning Foundation (2020KJRC0093).

## References

- [1] R. Arai and K. Murakami, "Hierarchical human motion recognition by using motion capture system," in *2018 International Workshop on Advanced Image Technology (IWAIT)*, pp. 1–4, Chiang Mai, Thailand, 7–10 January 2018.
- [2] I. Gajniyarov, I. Mikhailov, I. Starodubtsev et al., "The motion capture as behavior analyzing method of spontaneous motor activity in human infants," in *2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, pp. 681–684, Novosibirsk, Russia, 2019.
- [3] S. Sharma, S. Verma, M. Kumar, and L. Sharma, "Use of motion capture in 3D animation: motion capture systems,

- challenges, and recent trends,” in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 289–294, Faridabad, India, 2019.
- [4] Z. Yang, M. H. Rafiei, A. Hall et al., “A novel methodology for extracting and evaluating therapeutic movements in game-based motion capture rehabilitation systems,” *Journal of Medical Systems*, vol. 42, no. 12, p. 255, 2018.
  - [5] L. Noitom, “EB/OL,” 2020, <https://www.noitom.com.cn/>.
  - [6] K. Aberman, R. Wu, D. Lischinski, B. Chen, and D. Cohen-Or, “Learning character-agnostic motion for motion retargeting in 2D,” *ACM Transactions on Graphics*, vol. 38, no. 4, pp. 1–14, 2019.
  - [7] N. F. Troje, “Decomposing biological motion: a framework for analysis and synthesis of human gait patterns,” *Journal of Vision*, vol. 2, no. 5, pp. 371–387, 2002.
  - [8] A. Shapiro, Y. Cao, and P. Faloutsos, “Style components,” in *Proceedings of Graphics Interface 2006*, pp. 33–39, Québec, Canada, 2006.
  - [9] J. Wang and Bodenheimer, “Synthesis and evaluation of linear motion transitions,” *ACM Transactions on Graphics*, vol. 27, no. 1, pp. 1–15, 2008.
  - [10] T. Mukai and S. Kuriyama, *Geostatistical motion interpolation*, vol. 24, no. 3, 2005, ACM SIGGRAPH 2005 Papers, New York, NY, USA, 2005, Association for Computing Machinery.
  - [11] J. M. Wang, D. J. Fleet, and A. Hertzmann, “Gaussian process dynamical models for human motion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 283–298, 2008.
  - [12] B. Matthew and A. Hertzmann, “Style machines,” in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 183–192, USA, 2000.
  - [13] K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popovic, *Style-based inverse kinematics*, ACM SIGGRAPH 2004 Papers, New York, NY, USA, 2004.
  - [14] L. Kovar, M. Gleicher, and F. Pighin, “Motion graphs,” in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH’02*, pp. 473–482, New York, NY, USA, 2002.
  - [15] O. Arikan and D. Forsyth, “Interactive motion generation from examples,” in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’02 (2002)*, vol. 21no. 3, pp. 483–490, New York, NY, USA, 2002.
  - [16] J. Min and J. Chai, “Motion graphs++: a compact generative model for semantic motion analysis and synthesis,” in *ACM Transactions on Graphics*, vol. 31, no. 6pp. 1–12, Association for Computing Machinery, New York, NY, USA, 2012.
  - [17] Y. Lee, K. Wampler, G. Bernstein, J. Popović, and Z. Popović, *Motion fields for interactive character locomotion*, vol. 29, no. 6, 2010, ACM Transactions on Graphics (TOG), New York, NY, USA, 2010, Association for Computing Machinery.
  - [18] G. W. Taylor, G. E. Hinton, and S. T. Roweis, “Two distributed-state models for generating high-dimensional time series,” *Journal of Machine Learning Research*, vol. 12, pp. 1025–1068, 2011.
  - [19] R. Mittelman, B. Kuijpers, S. Savarese, and H. Lee, “Structured recurrent temporal restricted Boltzmann machines,” *Proceedings of Machine Learning Research*, P. X. Eric and J. Tony, Eds., , pp. 3620–3628, PMLR, Beijing, China, 2014.
  - [20] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, “Recurrent network models for human dynamics,” in *Proceedings of the IEEE International Conference on Computer Vision. 2015*, pp. 4346–4354, Santiago, Chile, 2015.
  - [21] Y. Du, W. Wang, and L. Wang, “Hierarchical recurrent neural network for skeleton based action recognition,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1110–1118, Boston, MA, 2015.
  - [22] D. Holden, J. Saito, and T. Komura, “A deep learning framework for character motion synthesis and editing,” *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–11, 2016.
  - [23] X. F. Zan, *Research on Editing and Reuse Technology of Human Motion Capture Data*, Beijing Jiaotong University, Beijing, China, 2019, A master’s degree.
  - [24] K. Aberman, Y. Weng, D. Lischinski, D. Cohen-Or, and B. Chen, “Unpaired motion style transfer from video to animation,” *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, 2020.
  - [25] Z. D. Yu, A. Andreas, S. Ariel, M. Moshe, and J. Eakta, “Adult 2child: motion style transfer using cycle GANs,” in *{MIG} ’20: Motion, Interaction and Games*, J. G. Stephen, S. Shinjiro, K. Ioannis, and B. Z. Victor, Eds., vol. 13, pp. 1–11, Virtual Event, SC, USA, 2020.
  - [26] S. J. Shin, S. C. You, H. Jeon et al., “Style transfer strategy for developing a generalizable deep learning application in digital pathology,” *Computer Methods and Programs in Biomedicine*, vol. 198, article 105815, 2021.
  - [27] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, Venice, Italy, 2017.
  - [28] Z. Yan, Z. Du, and D. Wu, “Ball interaction model for characteristics simulation of soft tissue,” *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/Journal of Computer-Aided Design and Computer Graphics*, vol. 26, no. 8, pp. 1346–1353, 2014.
  - [29] J. Lee and S. Y. Shin, “A hierarchical approach to interactive motion editing for human-like figures,” in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 39–48, ACM Press/Addison-Wesley Publishing Co., USA, 1999.
  - [30] S. Tak and H. S. Ko, “A physically-based motion retargeting filter,” *ACM Transactions on Graphics*, vol. 24, no. 1, pp. 98–117, 2005.
  - [31] K. Choi and H. Ko, “Online motion retargetting,” *The Journal of Visualization and Computer Animation*, vol. 11, no. 5, pp. 223–235, 2000.
  - [32] M. Gleicher, “Retargetting motion to new characters,” in *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH*, pp. 33–42, Budmerice Castle, Slovakia, 1998.
  - [33] Y. Zhang, L. Ye, J. Wang, and Q. Zhang, “Motion retargeting based on terminal effector constraints,” in *Proceedings of 2015 IEEE Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2015*, pp. 513–517, Chongqing, China, 2016.
  - [34] Y. Zhou, S. J. Li, H. S. Zhu, and X. P. Liu, “An all-purpose bidirectional recurrent autoencoder for retargeting of motion data represented by joint position,” *Journal of Computer-Aided Design & Computer Graphics*, vol. 32, no. 2, pp. 315–324+333, 2020.
  - [35] R. Villegas, J. Yang, D. Ceylan, and H. Lee, “Neural kinematic networks for unsupervised motion retargetting,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8639–8648, Salt Lake City, UT, 2018.



- [36] J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins, and N. S. Pollard, "Interactive control of avatars animated with human motion data," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 491–500, 2002.
- [37] J. Lee and K. H. Lee, "Precomputing avatar behavior from human motion data," *Graphical Models*, vol. 68, no. 2, pp. 158–174, 2006.
- [38] D. Hu, *Character Motion Synthesis and Style Transfer Based on Deep Learning and Spatio-Temporal Constraint*, Huaqiao University, Fujian, China, 2019, A master's degree.
- [39] X. Guo, S. Xu, W. Che, and X. Zhang, "Automatic motion generation based on path editing from motion capture data," in *Transactions on Edutainment IV*, pp. 91–104, Springer, Berlin, Heidelberg, 2010.