3D Animation Reconstruction Technology for Sports Skills that Integrates Motion Capture and Physical Simulation

Yu Jiang[1,*], Baosheng Hu[2], Baolai Bai[3]
[1]School of Design and Art, Shandong Huayu University of Technology, Dezhou 253034, China
[2]School of General Education, Shandong Huayu University of Technology, Dezhou 253034, China
[3]School of Information Engineering, Shandong Huayu University of Technology, Dezhou 253034, China
Corresponding author: Yu Jiang (e-mail: Jy13104155655@163.com).

ABSTRACT Motion capture data lack physical constraints such as ground reaction forces and joint torques, leading to penetration and imbalance in the resulting three-dimensional animations. This paper proposes a constraint-adversarial physics-aware reconstruction method for sports skill animation. Physical priors are embedded into motion features through graph convolutional encoding with a differential dynamics loss function. A policy network is then trained via Proximal Policy Optimization, whose reward function combines a Dynamic Time Warping-based style term with angular momentum conservation and contact point stability terms to generate joint driving torques. The poses are updated through implicit integration, and penetration errors are corrected by a Signed Distance Field closed-loop module. Experimental results demonstrate that the method achieves a mean joint position error of 2.3 cm, a physics validity score of 0.92, an average reconstruction time of 18 ms per frame, and a naturalness score of 4.7. The approach preserves motion style while producing physically plausible animation.

INDEX TERMS 3D Animation Reconstruction; Fusion of Motion Capture and Physical Simulation; Sports Skill Modeling; Motion Strategy Network; PPO Algorithm

I. INTRODUCTION

WITH the widespread application of 3D animation technology in sports science, film and television entertainment, and virtual training, generating animation sequences that retain the details of real athletes' movements while conforming to physical laws has become a key challenge [1]. Although traditional motion capture technology can accurately record the trajectory of human joints, it completely ignores physical constraints such as gravity, collision, and joint torque, resulting in obvious violations of physical laws in the animation in the virtual environment, such as limbs penetrating the ground and the inability to maintain posture in the air. Although pure physical simulation methods can ensure the self-consistency of kinematics and dynamics, they are difficult to reproduce the subtle style features and coordination patterns in real sports skills [2-3]. Therefore, integrating the advantages of motion capture data and physical simulation to achieve high-fidelity and physically reasonable 3D animation reconstruction of sports skills has important theoretical value and application prospects.

To address the above problems, this paper proposes a physical perception motion reconstruction method based on constraint adversarial. This method embeds physical priors into the motion feature encoding process by constructing a differential dynamic loss function, and uses a proximal strategy to optimize the training of a motion strategy network that integrates movement style and physical effectiveness. Furthermore, a signed distance field closed-loop correction module is designed to eliminate penetration error. The overall framework achieves end-to-end reconstruction from raw captured data to physically accurate animation frames, significantly enhancing the physical realism of the animation while maintaining the original motion style.

The organization of this paper is as follows: The introduction introduces the related work and related research situation; The methodology explains the specific implementation process of spatiotemporal feature encoding, policy network training, and implicit integration and collision correction; The experiments and results analysis, based on the sports skills dataset, prove the advantages of our method in motion fidelity, physics validity, and real-time performance; Finally, the conclusion summarizes the whole paper and makes future work.

II. RELATED WODR

On the problem of human motion reconstruction, most of the existing studies are focused on data-driven modeling, optimization of pose estimation, and motion sequence generation. Wang et al. utilized a dynamic dense graph convolutional network to model the spatiotemporal dependencies between human skeleton joints [4]; Pham designed a lightweight graph convolutional network to extract human joint structural features for the skeleton data action recognition [5]. While Starke et al. utilized the DeepPhase method, which casts a periodic autoencoder to learn a multidimensional motion phase space from unstructured human motion data [6]; Qin et al. employed a two-stage Transformer for the motion completion application, which is able to generate variable-length and relatively natural transition actions between context frames and target frames [7]. The above studies show that deep learning methods can effectively improve the feature representation, prediction, and completion capabilities of human motion sequences, but their optimization objectives are mostly focused on joint trajectory fitting, visual continuity, and motion naturalness, while not adequately considering physical factors such as ground reaction force, joint torque, angular momentum conservation, and contact stability. Therefore, in the reconstruction of three-dimensional animation of sports skills, relying solely on data-driven methods can still easily lead to problems such as foot slippage, body penetration, and posture imbalance.

To compensate for the shortcomings of pure data-driven methods in terms of physical realism, physical perception motion control and reinforcement learning methods have been gradually introduced into the reconstruction of 3D character animation. Won et al. proposed a physical character controller based on conditional variational autoencoders, which enables virtual characters to generate diverse motion behaviors similar to training motion data in a physical simulation environment [8]. Peng et al. further proposed a large-scale reusable adversarial skill embedding method, which trains physical characters to master multiple types of sports skills through adversarial imitation learning and unsupervised reinforcement learning [9]. For sports skill scenarios, Zhang et al. learned physical simulation hitting skills from tennis match videos, enabling virtual characters to complete sports actions with realistic motion styles under dynamic constraints [10]. However, existing methods still have shortcomings in the unified modeling of motion capture data and dynamic constraints, contact collision feedback correction, and the preservation of sports skill action styles. Based on this, this paper introduces a constraint adversarial physical perception motion reconstruction method, which combines graph convolutional spatiotemporal feature extraction, PPO motion strategy optimization, differential dynamic loss, and signed distance field collision feedback to achieve high-fidelity, physically consistent 3D animation reconstruction of sports skill actions.

III. METHOD

A. MOTION FEATURE ENCODING INCORPORATING PHYSICAL CONSTRAINTS

The raw motion capture data is input as a sequence of joint trajectories. Each time frame contains the 3D position vectors $p_j \in \mathbb{R}^3$ and rotation angle quaternions $q_j \in \mathbb{R}^4$ of each joint in the human skeleton, where $j = 1, \ldots, J$ are joint indices. This paper constructs a spatiotemporal feature encoder, outputting a latent motion feature tensor $Z \in \mathbb{R}^{T\times J\times D}$ embedded with physical priors, where T is the sequence length and D is the feature dimension. The encoder first performs graph convolution on the spatial structure of each frame. An undirected graph $G = (V, E)$ of the human skeleton is defined, where the node set V corresponds to J joints, and the edge set E represents the skeletal connections. The graph convolutional layer takes the joint features $H_t^{(0)} = [p_{1,t}; q_{1,t}; \ldots; p_{J,t}; q_{J,t}] \in \mathbb{R}^{J\times 7}$ of frame t as input and performs spatial feature aggregation using a graph Laplacian operator normalized by the adjacency matrix.

$$H_t^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H_t^{(l)}W^{(l)}\right) \tag{1}$$

In formula (1), $\tilde{A} = A + I_J$ is the adjacency matrix of added self-loop, A is the adjacency matrix of the skeleton graph, I_J is the identity matrix; $\tilde{D}$ is the degree matrix of $\tilde{A}$; $W^{(l)}$ is the learnable weight matrix of the l-th layer; $\sigma(\cdot)$ is the ReLU activation function; $H_t^{(l)}$ is the node feature matrix output by the l-th layer, and $H_t^{(L)}$ is obtained after convolution of the L layer graph. Then the sequence $\{H_1^{(L)}, \ldots, H_T^{(L)}\}$ in the time dimension is input into the one-dimensional sequential convolution network, the convolution kernel slides along the time axis, and the feature tensor $F \in \mathbb{R}^{T\times J\times D}$ is output.

To embed physical constraints into the feature representation, a differential dynamics loss function is constructed. The joint acceleration a_t^{mocap} is defined as the second-order difference of the joint position $p_{j,t}$ with respect to time in the capture data. Based on the physical simulation environment, the physically compliant acceleration a_t^{phy} is obtained through inverse dynamics calculation using the ground reaction force $f_t^{\text{grf}} \in \mathbb{R}^3$ and the joint torque $\tau_t \in \mathbb{R}^J$. The differential dynamics loss function is in the form of:

$$\mathcal{L}_{\text{dyn}} = \frac{1}{T}\sum_{t=1}^{T}\left\|a_t^{\text{mocap}} - a_t^{\text{phy}}\right\|_2^2 \tag{2}$$

In equation (2), a_t^{mocap} and a_t^{phy} are both $3J$-dimensional vectors, representing the acceleration of all joints in frame t, respectively; $\|\cdot\|_2^2$ is the square of the Euclidean distance. This loss function propagates back through the training process, forcing the feature tensor Z output by the spatiotemporal feature encoder to contain a motion pattern consistent with physical dynamics. Z is ultimately passed as input to the subsequent motion policy network [11-12].

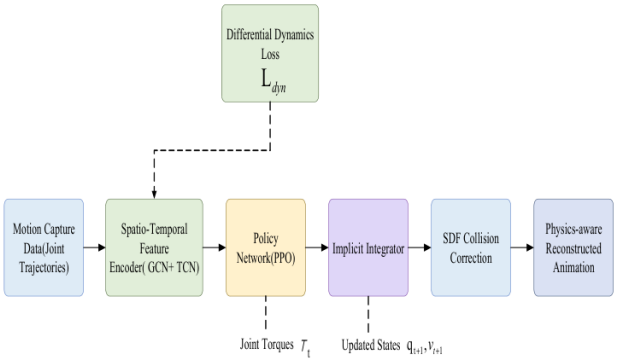

FIGURE 1. Overall framework for 3D animation reconstruction of sports skills based on physical perception

Fig. 1 shows the overall framework of the proposed method, indicating the introduction of the differential dynamic loss function $\mathcal{L}_{\text{dyn}}$ and the flow from the encoder output to the policy network.

B. PHYSICS-AWARE MOTION POLICY NETWORK BASED ON PROXIMAL POLICY OPTIMIZATION

The latent motion feature tensor $Z \in \mathbb{R}^{T\times J\times D}$ output by the spatiotemporal feature encoder is concatenated with the skeleton state vector s_t in the current simulation environment and used as the input to the motion policy network. s_t contains the joint angle $\theta_t \in \mathbb{R}^J$, joint angular velocity $\dot{\theta}_t \in \mathbb{R}^J$, and ground reaction force $f_t^{\text{grf}} \in \mathbb{R}^3$ of frame t, with a total dimension of $2J+3$. The policy network $\pi_\phi(u_t|s_t, Z_t)$ is trained using Proximal Policy Optimization (PPO) and outputs the action u_t, i.e., the driving torque $\tau_t \in \mathbb{R}^J$ of each joint.

The objective function of PPO prunes the probability ratio in each policy update to avoid excessively large update steps. The probability ratio is defined as $r_t(\phi) = \frac{\pi_\phi(u_t|s_t,Z_t)}{\pi_{\phi_{\text{old}}}(u_t|s_t,Z_t)}$, and the pruned loss function is:

$$\mathcal{L}^{\text{CLIP}}(\phi) = E_t\left[\min\left(r_t(\phi)\hat{A}_t, \text{clip}(r_t(\phi), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right] \tag{3}$$

In formula (3), $\hat{A}_t$ is the estimated value of the advantage function at time t, ϵ is the clipping hyperparameter (set to 0.2), the $\text{clip}(\cdot)$ function restricts the ratio within the interval $[1-\epsilon, 1+\epsilon]$, and E_t represents the expected value of the time step.

The reward function R_t is composed of a weighted sum of the action style term R_t^{style} and the physical effectiveness term R_t^{phys}. The weight coefficients $\alpha = 0.6$ and $\beta = 0.4$ are determined by grid search, as shown in formula (4):

$$R_t = \alpha R_t^{\text{style}} + \beta R_t^{\text{phys}} \tag{4}$$

The motion style term measures the temporal similarity between the current simulated pose and the motion capture data. The Dynamic Time Warping (DTW) distance, $D_{\text{DTW}}(p^{\text{sim}}, p^{\text{mocap}})$, calculates the minimum alignment cost between the simulated joint position sequence p^{sim} and the capture sequence p^{mocap}. The DTW distance is normalized and mapped to the reward interval $[-1, 0]$.

$$R_t^{\text{style}} = -\frac{D_{\text{DTW}}\left(p_{1:t}^{\text{sim}}, p_{1:t}^{\text{mocap}}\right)}{D_{\max}} \tag{5}$$

In formula (5), $p_{1:t}^{\text{sim}}$ and $p_{1:t}^{\text{mocap}}$ are the simulated and captured joint position sequences from the initial time to time t, respectively, and $D_{\max}$ is the normalization factor for the maximum DTW distance in the dataset [13].

The physical validity term R_t^{phys} is further decomposed into an angular momentum conservation deviation term and a contact point stability term. The angular momentum conservation deviation ΔL_t is defined as the difference norm between the total angular momentum L_t of the current frame and the angular momentum L_0 of the initial frame. The contact point stability term calculates the slip distance δ_t^{slip} between the foot and the ground contact point between two consecutive frames. The two terms are combined as follows:

$$R_t^{\text{phys}} = -\lambda_1 \left\|\Delta L_t\right\|_2 - \lambda_2 \delta_t^{\text{slip}} \tag{6}$$

In formula (6), $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$ are balance coefficients; $\Delta L_t = L_t - L_0$, angular momentum $L_t = \sum_j r_{j,t} \times (m_j v_{j,t})$, $r_{j,t}$, m_j, $v_{j,t}$ are the position vector, mass and linear velocity of the j-th joint at time t, respectively; $\delta_t^{\text{slip}} = \sum_{c\in C} \|x_{c,t} - x_{c,t-1}\|_2$, C is the set of contact points, and $x_{c,t}$ is the position of the contact point in frame t.

The policy network is trained using the PPO algorithm. In each round, T steps of trajectory are sampled, the advantage function $\hat{A}_t$ and reward R_t are calculated, and the policy network parameters ϕ are updated. Simultaneously, the value network is updated to reduce value function estimation errors. After training convergence, the policy network outputs the joint driving torque τ_t, which is passed to the implicit integrator for the next stage of processing.

C. IMPLICIT INTEGRAL AND SIGNED DISTANCE FIELD LOOP CORRECTION

The joint driving torque $\tau_t \in \mathbb{R}^J$ output by the motion policy network is fed into the implicit integrator. The skeletal model consists of J rigid body segments, each with mass m_j, inertia tensor $I_j \in \mathbb{R}^{3\times 3}$, and centroid position. Given the generalized coordinates $q_t = [\theta_t; x_t^{\text{root}}]$ of the current frame, where $\theta_t \in \mathbb{R}^J$ are joint angles and $x_t^{\text{root}} \in \mathbb{R}^3$ are root node positions. The implicit integrator uses a backward Euler scheme, updating both velocity and position simultaneously.

$$v_{t+1} = v_t + \Delta t\, M^{-1}\left(\tau_t - C(q_t, v_t)\right) \tag{7}$$

$$q_{t+1} = q_t + \Delta t\, v_{t+1} \tag{8}$$

In formulas (7) and (8), v_t is the generalized velocity vector (including joint angular velocity and root node linear velocity), Δt is the simulation time step (set to 0.02 seconds), M is the generalized mass matrix, and $C(q_t, v_t)$ is the combination of Coriolis force and gravity term. Formula (7) is an implicit equation. Since v_{t+1} appears on both sides of the equation, it is solved by Newton's iteration method. After the iteration converges, the updated generalized coordinate q_{t+1} is obtained, which corresponds to the position and orientation of each bone segment in three-dimensional space [14].

The updated skeleton pose may penetrate environmental surfaces (ground, obstacles). To correct for penetration, a signed distance field (SDF) is constructed. The SDF is a scalar function $\Phi(x)$ defined in three-dimensional space. For any point $x \in \mathbb{R}^3$ in space, it outputs the signed distance from that point to the nearest environmental surface: positive for the outside, negative for the inside, and zero for the surface. For the set of surface sampling points $P_j = \{x_{j,k} \mid k = 1, \ldots, K\}$ on each skeleton segment, the signed distance $\Phi(x_{j,k})$ of each sampling point is calculated. The penetration depth is defined as the maximum absolute value of all negative distances, as shown in formula (9):

$$d_{\text{pen}} = \max_{j,k} \left(-\Phi(x_{j,k}), 0\right) \tag{9}$$

When $d_{\text{pen}} > 0$, the position needs to be corrected. A differentiable penetration depth gradient is introduced to backpropagate the penetration error to the bone segment position. The collision loss function is defined as follows:

$$\mathcal{L}_{\text{coll}} = \sum_{j=1}^{J} \sum_{k=1}^{K} \max\left(-\Phi(x_{j,k}), 0\right)^2 \tag{10}$$

In formula (10), $x_{j,k}$ depends on the generalized coordinate q_{t+1}, therefore $\mathcal{L}_{\text{coll}}$ is differentiable with respect to q_{t+1}. Calculate the gradient $\nabla_q \mathcal{L}_{\text{coll}}$, and correct the generalized coordinates along the gradient descent direction:

$$q_{t+1}^{\text{corr}} = q_{t+1} - \eta \nabla_q \mathcal{L}_{\text{coll}} \tag{11}$$

In formula (11), $\eta = 0.1$ is the correction step size. This correction process is performed iteratively until $d_{\text{pen}} \leq 10^{-5}$ meters. The final output is the non-penetrating generalized coordinate q_{t+1}^{corr}, which generates the complete animation pose and position of frame $t+1$ as the initial state for the next time step.

Table 1 shows the average penetration depth (in millimeters) of different bone segments in the three stages before correction, after a single correction, and after iterative convergence, quantifying the effectiveness of the closed-loop correction.

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

The experiment used publicly available sports skill motion capture datasets, including basketball, gymnastics, and track and field. Each category contained 10 standard motion sequences, for a total of 15,000 frames, with a sampling rate of 120 frames per second. Each frame provided the 3D positions and rotational quaternions of 23 joints of the human skeleton, and the skeleton topology conformed to a standard biomechanical model. The ground model of the physical simulation environment was set as a horizontal rigid plane, with a gravitational acceleration of 9.8 m/s^2. The mass distribution of the skeletal segments was assigned based on anthropometry databases, and the contact points were set as five key nodes between the soles of the feet and the ground.

TABLE 1. Average Penetration Depth of Different Bone Segments During SDF Closed-Loop Correction Process

Body segment	Before correction	After single correction	After iterative convergence
Torso	3.2	0.8	0.02
Thigh	4.5	1.2	0.03
Calf	5.1	1.5	0.04
Foot	6.3	2.0	0.05

B. MOTION FIDELITY ANALYSIS

To assess the overall fidelity of the motion, the error distribution was derived from three dimensions: joint space, motion type and time history, for which the average joint position error was 2.3 cm. In joint space, the average error of 23 joints were 1.5 cm to 3.4 cm with more error at the distal joint (wrist and ankle) (3.2 cm for wrist and 3.4 cm for ankle), moderate error at the proximal joint (shoulder and hip) (2.4 cm to 2.7 cm), and least error at the core trunk joint (thoracic and lumbar vertebrae) (1.5 cm to 1.8 cm). The errors for the ten common motion sequences are: basketball layup 2.6 cm, basketball jump shot 2.4 cm, gymnastics front handspring 2.9 cm, gymnastics backflip 3.1 cm, track and field 100-meter start 1.9 cm, track and field long jump 2.5 cm, track and field javelin throw 2.2 cm, gymnastics side handspring 2.4 cm, basketball dribbling 2.1 cm, and track and field hurdles 2.3 cm. From the time history perspective, the peak error is at the 28th frame, and the maximum error is 3.4 cm after standardizing the motion to 100 frames. The error distribution in the three dimensions is shown in Fig. 2, which indicates that the error in the initial frame (first 15 frames) is less than 1.6cm (less than 1 percent of the length of the initial frame), in the final frame (last 15 frames) is less than 1.6cm (less than 1 percent of the length of the final frame), and the standard deviation of the error fluctuation in the middle frame (frames 30-70) is 0.86cm.

The results validate the effectiveness of the proposed differential dynamics loss function and DTW reward term in preserving original motion details: low error values indicate a high degree of consistency between the simulated and captured postures in the spatial trajectory, while the gradient changes in error at different motion stages reflect the adaptive adjustment capability of the policy network to

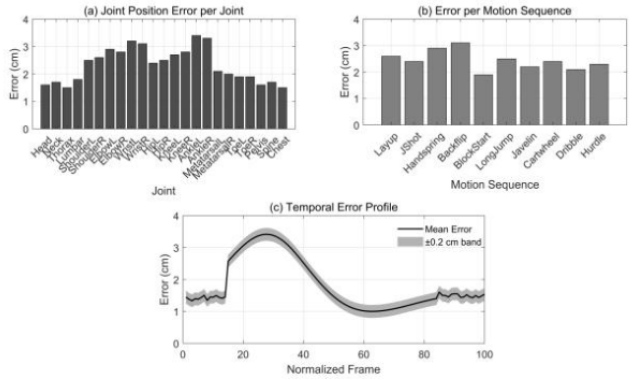

FIGURE 2. Motion fidelity analysis. (a) Joint position errors. (b) Motion sequence errors. (c) Time history error curves

physical driving forces.

C. VALIDATION OF PHYSICAL EFFECTIVENESS

Physical effectiveness is jointly evaluated using three sub-indicators: the number of penetration events (each frame with a penetration depth exceeding 1 mm is counted as one event), the mean angular momentum conservation deviation (the average of the norm of total angular momentum change across all frames, in kg·m^2/s), and the mean contact point slip distance (the average displacement of the contact point between consecutive frames, in mm). The physical effectiveness score is a weighted sum of the normalized three sub-indicators, ranging from 0 to 1. The overall score for our method is 0.92. Table 2 provides detailed values for the physical effectiveness sub-indicators of our method in basketball, gymnastics, and track and field, and compares them with pure motion capture redirection, pure physics simulation, existing reinforcement learning fusion methods, and the current state-of-the-art method AMP (Adversarial Motion Prior).

TABLE 2. Comparison of Physical Effectiveness Indicators of Different Methods

Method	Penetration events (per frame)	Mean angular momentum deviation (kg·m^2/s)	Mean contact point slip distance (mm)	Physical validity score
Pure motion capture retargeting	0.87	5.62	12.4	0.31
Pure physics simulation	0	0.31	1.2	0.95
Existing reinforcement learning fusion method	0.12	1.85	3.8	0.76
AMP (Adversarial Motion Prior)	0.04	1.12	2.9	0.84
Proposed method	0.01	0.52	1.8	0.92

In Table 2, the number of penetration events was 0.01 times/frame, corresponding to one penetration depth exceeding 1 mm every 100 frames, and this penetration was immediately eliminated after closed-loop correction. The average angular momentum conservation deviation was 0.52 kg·m^2/s, indicating that the change in total angular momentum was effectively constrained, and no uncontrollable tumbling occurred in jumping and rotating movements. The average contact point slip distance was 1.8 mm, and the relative sliding of the foot to the ground during the support phase was less than 2 mm, which is consistent with the stable range of foot-ground contact in biomechanics. The sub-index decomposition of the three types of sports showed that the angular momentum deviation was slightly higher in gymnastics (0.68 kg·m^2/s), the number of penetration events was slightly higher in basketball (0.02 times/frame), and the slip distance was slightly smaller in track and field (1.5 mm), reflecting the adaptive adjustment under different dynamic characteristics. Compared with pure physics simulation, the physical effectiveness score of the proposed method was slightly lower by 0.03, but pure physics simulation completely abandoned the motion style fidelity (its motion fidelity error was as high as 7.8 cm, far higher than the 2.3 cm of the proposed method). Compared to existing reinforcement learning fusion methods, the proposed method achieves a physics-effective score 0.16 higher; compared to the AMP method, it achieves a physics-effective score 0.08 higher, reduces penetration events by 75%, and lowers angular momentum bias by 54%, demonstrating the advantages of the signed range field closed-loop correction module in reducing penetration and controlling slip.

D. REAL-TIME PERFORMANCE AND SUBJECTIVE EVALUATION

The proposed method achieves an average single-frame reconstruction time of 18 milliseconds on a specified hardware environment (Intel Core i9-13900K CPU, NVIDIA GeForce RTX 4090 GPU, 64GB RAM), corresponding to a processing rate of approximately 55.6 frames per second, meeting the requirements for real-time animation generation (typically requiring $\geq$30 frames per second). The time consumption of each module was measured separately: the spatiotemporal feature encoding stage (graph convolution and temporal convolution) averaged 3.2 milliseconds; policy network inference (outputting joint driving torque) averaged 1.5 milliseconds; implicit integration and Newton iteration averaged 6.8 milliseconds; and signed distance field query and gradient correction iteration averaged 6.5 milliseconds. The SDF correction module accounts for 36.1% of the total time, representing the main computational bottleneck, but it is crucial for improving physical effectiveness.

For subjective evaluation, five referees with national first-level athlete qualifications were invited to independently view the animation generated by the proposed method, using a five-point Likert scale (1 point = very unnatural, 5 points = very natural) to score the naturalness of the movements. The final average score was 4.7, with all judges' scores ranging from 4.5 to 5.0. Judges' feedback indicated that the animation's landing cushioning, torso twisting, and limb coordination closely resembled the performance of a real

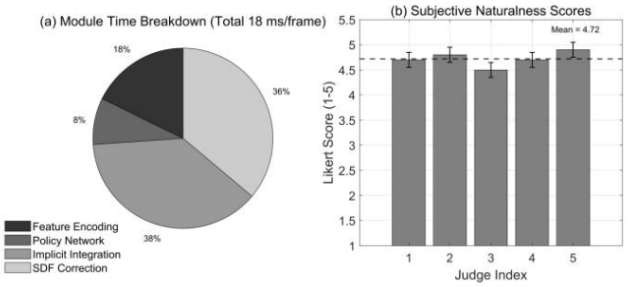

FIGURE 3. Real-time performance and subjective evaluation. (a) Module elapsed time percentage. (b) Subjective rating distribution

athlete, and no obvious penetration or sliding was observed. Fig. 3 shows the time allocation for each module and the distribution of subjective scores.

V. CONCLUSION

This paper presents an approach to reconstructing 3D animation of sports skills through combining motion capture and physical simulation. The method effectively tackles the conflict between physical distortion of pure motion capture data and lack of style fidelity of pure simulation by using graph convolutional feature encoding, near-end policy optimization training and signed distance field closed-loop correction. The results of experiments indicate that this approach can greatly enhance the physical plausibility of the animation while retaining motion details. The current approaches do not have the flexibility to adapt to complex environmental interactions (e.g., non-rigid ground and multi-person collaboration) and the SDF correction module is computationally expensive. For more complex sports scenarios and real-time interactive applications, future work will focus on dynamic collision prediction and lightweight inference approaches leveraging graph neural network to further facilitate the sports understanding and prediction process.

REFERENCES

[1] R. Rekik, S. Wuhrer, L. Hoyet, K. Zibrek, and A. H. Olivier, "A Survey on Realistic Virtual Human Animations: Definitions, Features and Evaluations," Computer Graphics Forum, vol. 43, no. 2, pp. e15064:1–e15064:21, 2024. DOI: 10.1111/cgf.15064.

[2] A. Kwiatkowski, E. Alvarado, V. Kalogeiton, C. K. Liu, J. Pettre, M. van de Panne, and M. P. Cani, "A Survey on Reinforcement Learning Methods in Character Animation," Computer Graphics Forum, vol. 41, no. 2, pp. 613–639, 2022. DOI: 10.1111/cgf.14504.

[3] T. J. Dobos, R. W. G. Bench, C. D. McKinnon, A. Brady, K. J. Boddy, M. W. R. Holmes, and M. W. L. Sonne, "Validation of pitchAI markerless motion capture using marker-based 3D motion capture," Sports Biomechanics, vol. 24, no. 3, pp. 587–607, 2025. DOI: 10.1080/14763141.2022.2137425.

[4] X. Wang, W. Zhang, C. Wang, Y. Gao, and M. Liu, "Dynamic Dense Graph Convolutional Network for Skeleton-Based Human Motion Prediction," IEEE Transactions on Image Processing, vol. 33, pp. 1–15, 2024. DOI: 10.1109/TIP.2023.3334954.

[5] D. T. Pham, Q. T. Pham, T. T. Nguyen, T. L. Le, and H. Vu, "A lightweight graph convolutional network for skeleton-based action recognition," Multimedia Tools and Applications, vol. 82, no. 2, pp. 3055–3079, 2023. DOI: 10.1007/s11042-022-13298-w.

[6] S. Starke, I. Mason, and T. Komura, "Deepphase: Periodic autoencoders for learning motion phase manifolds," ACM Transactions on Graphics (ToG), vol. 41, no. 4, pp. 1–13, 2022. DOI: 10.1145/3528223.3530178.

[7] J. Qin, Y. Zheng, and K. Zhou, "Motion In-Betweening via Two-Stage Transformers," ACM Transactions on Graphics, vol. 41, no. 6, pp. 184:1–184:16, 2022. DOI: 10.1145/3550454.3555454.

[8] J. Won, D. Gopinath, and J. Hodgins, "Physics-based character controllers using conditional vaes," ACM Transactions on Graphics (TOG), vol. 41, no. 4, pp. 1–12, 2022. DOI: 10.1145/3528223.3530067.

[9] X. B. Peng, Y. Guo, L. Halper, et al., "Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters," ACM Transactions On Graphics (TOG), vol. 41, no. 4, pp. 1–17, 2022. DOI: 10.1145/3528223.3530110.

[10] H. Zhang, Y. Yuan, V. Makoviychuk, Y. Guo, S. Fidler, X. B. Peng, and K. Fatahalian, "Learning Physically Simulated Tennis Skills from Broadcast Videos," ACM Transactions on Graphics, vol. 42, no. 4, pp. 95:1–95:14, 2023. DOI: 10.1145/3592408.

[11] B. Filtjens, B. Vanrumste, and P. Slaets, "Skeleton-Based Action Segmentation with Multi-Stage Spatial-Temporal Graph Convolutional Neural Networks," IEEE Transactions on Emerging Topics in Computing, vol. 12, no. 1, pp. 202–212, 2024. DOI: 10.1109/TETC.2022.3230912.

[12] Q. Men, H. P. H. Shum, E. S. L. Ho, and H. Leung, "GAN-Based Reactive Motion Synthesis with Class-Aware Discriminators for Human-Human Interaction," Computers & Graphics, vol. 102, no. 1, pp. 634–645, 2022. DOI: 10.1016/j.cag.2021.09.014.

[13] X. Zhang, Z. Chang, Q. Men, and H. P. H. Shum, "Physics-Based Motion Tracking of Contact-Rich Interacting Characters," Computer Graphics Forum, vol. 45, no. 2, pp. e70336:1–e70336:9, 2026. DOI: 10.1111/cgf.70336.

[14] J. Wang, T. Zhang, Q. Zhang, et al., "Implicit swept volume sdf: Enabling continuous collision-free trajectory generation for arbitrary shapes," ACM Transactions on Graphics (TOG), vol. 43, no. 4, pp. 1–14, 2024. DOI: 10.1145/3658181.