Sports Training Posture Simulation Based On Optical Imaging Sensor And Motion Trajectory Prediction Algorithm

Qu Meili[1], Chang Ning[1], Xing Enqian[1], and Gao Xuan[2*]

[1] *College of Physical Education, Yanching Institute of Technology, Langfang, Hebei 065201, China*
[2] *P.E. Department, Hebei University of Water Resources and Electric Engineering, Cangzhou, Hebei, 061000, China*
[*] *Corresponding author. E-mail: xuangao1230@163.com*

Received Nov. 3, 2025; accepted Jan. 3, 2026

In the evolving landscape of intelligent computing and digital modeling, simulating complex human activities requires a synthesis of physical accuracy and data-driven generalization, particularly in fields such as athletic performance analysis. Within the broader context of trustworthy intelligent systems and applied AI frameworks, this work focuses on the computational modeling of structured human motion through an integrated architecture that simulates and adapts physical postures in dynamic environments. Traditional approaches to motion simulation often lack adaptability across diverse morphological profiles and task semantics, resulting in reduced biomechanical plausibility and limited contextual alignment. These methods frequently overlook higher-order kinematic constraints, rely excessively on frame-wise prediction, and exhibit poor cross-domain generalization. To address these challenges, a hybrid framework is proposed, comprising a Kinematic-Aware Neural Projection Network (KANet) and a Biomechanically-Constrained Latent Adaptation (BCLA) module. KANet preserves temporal and anatomical dependencies via a dual-stream attention mechanism and graph-based encoding, enabling precise motion sequence synthesis within a symbolic latent space. By embedding structural priors and joint-specific constraints directly into the modeling process, it enhances interpretability and physical fidelity. BCLA complements this by introducing a contextual adaptation layer that integrates biomechanical heuristics and soft constraints to couple motion trajectories with personalized physical parameters and task conditions. Expressing adaptation as a residual transformation across the latent motion manifold, BCLA enhances robustness and zero-shot generalizability. Experimental results demonstrate significant improvements in biomechanical plausibility, motion smoothness, and contextual adaptability compared to state-of-the-art generative models.

Keywords: Symbolic Modeling, Biomechanical Constraint, Motion Adaptation, Generative Learning, Intelligent Systems

http://dx.doi.org/10.6180/jase.202610_33.056

1. Introduction

In recent years, the quest for optimized sport performance and injury minimization has motivated the development of smart sport training systems. Among them, posture simulation with optical imaging sensors in conjunction with motion trajectory forecasting algorithms has remained highly prominent [1, 2]. This activity is equally essential for the improvement of athletes' technique and biomechanical efficacy, in the role of informing real-time feedback and long-term training maximization [3]. Classical training depends heavily on subjective observation, which remains limited in accuracy and reproducibility. Conversely, automated posture simulation offers objective and quantitative feedback in support of personalized training approaches [4]. Additionally, the incorporation of motion forecasting within such systems enables the prediction of forthcoming motion patterns and the implementation of proactive cor-

rections and adaptive training [5]. These systems prove particularly valuable in situations in which the presence of real-time surveillance and minimal equipment disturbance is called for, such as in the case of the sport performance laboratory and training camps under field conditions.

To compensate for the shortcomings of manual observation and basic motion recording, the earliest approaches drew on symbolic knowledge representation and knowledge-based systems [6]. These systems endeavored to represent human posture and movement in terms of pre-defined biomechanical rules and expert-designed ontologies. Human joints and skeletal structures were represented in terms of kinematic chains, for example, and rules for motion formulated in terms of logical constraints [7]. Although such methods afforded interpretability and fairly facile integration with the expert knowledge representation base, they proved fragile and non-adaptive in the face of novel or complex motion patterns [8]. That they drew on manually authored models constrained scalability and made it difficult to accommodate the variety inherent in various sport disciplines and individual performers [9]. To eliminate the shortcomings of fixed-rule modeling, the following approaches started integrating data-driven aspects in order to enhance the robustness and flexibility.

To enhance flexibility and minimize programmed rule dependence, data-oriented and machine learning techniques emerged as a promising alternative [10]. The methods used large annotated databases of wearable sensors and optical imaging systems to learn statistical patterns to detect posture and predict motion trajectories [11]. Support vector machines, random forests, and hidden Markov models were some of the most used techniques to encode the complex movement dynamics [12]. This evolution enabled the systems to learn through empirical examples and generalize better across different individuals and movement patterns. Nevertheless, despite better performance, the models typically required large-scale feature engineering and could not cope with high-dimensional input data [13]. They also could not model long-term temporal dependencies typical of human motion and optimized trajectory forecasts in dynamic or multi-step activities. The shortcomings thus called for more research in deep learning-based methods that could learn time and hierarchical features directly from raw sensor observations.

To avoid the limitations of conventional machine learning methods, pre-trained models and deep learning have become widely used in motion forecasting and simulation of posture scenarios [14]. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), e.g., long short-term memory (LSTM) networks, have proved effective in learning space and time features from sequential motion streams [15]. More recently, transformer architectures and self-supervised learning techniques have made it worthwhile to tap large motion datasets without the necessity of labelling everything. This facilitates systems to pre-train with large movement patterns and fine-tune for sport-specific scenarios [16]. These models shine in capturing intricate, non-linear motion trajectories and provide better generalizability across highly diverse environments and sensor configurations. Nevertheless, deep learning methods consume much computational power and tend to be perceived as black-box systems with limited interpretability and diminished confidence in safety-critical scenarios such as impact sports. Moreover, real-time execution is also compromised by latency and hardware constraints in handheld devices. The limitations call for superior methods inheriting the interpretability of symbolic models, the flexibility of data-driven methods, and the strength of deep neural architectures.

According to the aforementioned shortcomings in conventional methods, such as the stiffness of symbolic models, the dependency of machine learning methods on features, and the resource consumption of deep networks, this paper presents an innovative sports training simulation method by combining optical imaging sensor information with an advanced motion trajectory prediction algorithm. The method remedies major weaknesses in the current research by employing a hybrid framework that unifies interpretable biomechanical models with deep temporal dynamics modeling. The system is capable of accommodating real-time inference with strong accuracy and meanwhile allowing adaptability to various sports scenarios and subject-specific motion features. By extensive experimentation and validation, the method enhances the accuracy and robustness of the prediction and facilitates interactive feedback training sessions. The framework provides us with a promising future direction in smart sports training based on the unification of the merits of old methods and the compensation of their shortcomings in coherent integration and optimization.

- A biomechanically inspired modeling framework with the embedding of deep temporal sequence learning to generate an accurate and interpretable simulation of posture is proposed.
- The system acquires robust adaptability and performance in several sports disciplines and training scenarios with real-time processing and generalization.
- Experiment results confirm superior performance of the prediction accuracy and quality of real-time feed-

back in comparison with state-of-the-art methods, and successful motion error minimization.

Optical imaging sensors have also become an essential tool in the accurate capture of motion data in the domain of sports training applications [17]. Optical imaging sensors, especially the structured light or time-of-flight (ToF)-based sensors, offer the potential for high-resolution imaging with non-invasive and markerless tracking technologies without disrupting athlete performance [18]. One major development has been the incorporation of depth-sensing technologies in commercialized sensors like Microsoft Kinect, Intel RealSense, and other RGB-D cameras. These sensors merge the power of RGB imaging with depth vision such that three-dimensional skeletal models can be generated in real time [19]. The latest studies have emphasized increasing the accuracy in the determination of the joints in the image sequence without the explicit detection of joints by deep learning methods like convolutional pose machines (CPMs) and graph convolutional networks (GCNs), which deduce the positions of joints directly from the spatial-temporal patterns in the image sequence. The resultant models do not just enhance the accuracy in determining the joints but also turn out to be more resilient to occlusions and lighting conditions that prove typical in the domain of sports [20]. Further, multi-view configurations of optical sensors become increasingly used in order to compensate for the limitations of single-view configurations and allow full 360-degree motion analysis [21]. Calibration techniques and sensor fusion methods become necessary in such configurations in order to ensure geometric consistency and minimize spatial-temporal error. A major concern remains the trade-off between spatial resolution and processing latency in the applications with real-time feedback, wherein correction necessitates immediate action in order to allow effective coaching [22]. Researchers also explored the promise of incorporating inertial measurement units (IMUs) in order to complement optical sensors and thus create hybrid systems that leverage the strengths of both modalities. Such systems can track fine movement and correct for error due to drift or occlusion [23].

The applications of optical sensor-based motion capture extend beyond biomechanics in order to include performance analysis, injury prevention, and adaptive training regimens based on the profile of the athlete in terms of biomechanics. As computational models become ever more sophisticated, the integration between optical sensors with motion interpretation involving the use of AI is expected to lead to ever-smarter and personalized sports training systems.

Posture detection in the area of sports training involves the analysis of intricate, high-dimensional motion data that is difficult to process with conventional computer vision algorithms [24]. The introduction of deep learning techniques, such as the convolutional neural network (CNN), recurrent neural network (RNN), and transformer models, has greatly facilitated the automated learning of spatial and temporal features meaningful to the task of posture classification and detection of abnormal motion. These paradigms may be trained in an end-to-end mode on large-scale annotated datasets reflecting the variability of the posture between different people, sports activities, and states of performance. CNN models prove themselves best at the extraction of local spatial features at the level of separate frames or 2D maps of joints, whereas the RNN and long short-term memory (LSTM) networks extract time relations in the sequence of postures, allowing dynamic analysis of motion [25].

More recently, spatio-temporal graph convolutional networks (ST-GCNs) have gained prominence for their ability to model the skeletal structure as a dynamic graph, where joints are nodes and bones are edges. This representation aligns naturally with human biomechanics and allows the network to learn meaningful patterns of movement directly from graph-structured data. The application of attention mechanisms, especially within transformer-based models, has further enhanced performance by enabling the selective focus on key joints or motion segments that are critical for specific postural assessments [26].

A RRT algorithm-based method, proposed by Karaman et al. [27] could promptly find a feasible path and then improve the result if additional computation time was allocated. The authors improved RRT by combining heuristics and branch-and-bound pruning of the search tree for their algorithm, which was tested using Monte Carlo simulation and outperformed the basic RRT method in terms of efficiency in both simulated and realistic experiments involving an outdoor robotic vehicle. To address the scarcity of labeled training data in certain sports, techniques such as data augmentation, transfer learning, and semi-supervised learning are widely employed [28].

Researchers have also explored multi-task learning approaches where models simultaneously perform posture classification, anomaly detection, and trajectory prediction, leading to improved generalization and reduced overfitting. Despite these advancements, challenges persist in handling occlusions, inter-athlete variability, and domain adaptation across different sports or sensor modalities. Integration with feedback systems, such as augmented reality interfaces or auditory cues, further enhances the usability of deep learning-based posture recognition systems in real-

world training settings [29]. These systems enable coaches and athletes to receive immediate, data-driven feedback, thereby accelerating skill acquisition and reducing the risk of injury due to poor form or repetitive strain.

Predicting future motion trajectories of athletes is a critical component in intelligent sports training systems, as it enables proactive coaching, performance optimization, and injury prevention through anticipatory guidance [30]. Recent developments have centered on data-driven trajectory forecasting algorithms fueled by machine learning and deep generative models [31].

Recurrent neural networks (RNNs), and more recently sequence-to-sequence (Seq2Seq) models, were used to learn time dependencies from previous motion sequences and generate plausible future poses. These architectures have been supplemented with the incorporation of attention mechanisms, allowing the network to attend to key temporal frames or spatial features driving future movement [32]. Generative adversarial networks (GANs) have also been used to generate diverse and realistic samples of trajectories, recovering the inherent uncertainty in human motion forecasting. Transformers with their long-range modeling of temporality are increasingly used for multi-step trajectory forecasting, in particular, coupled with skeletal graph representations, which maintain the connectivity of joints and spatial constraints. Multi-modal methods enrich the forecasting process with the incorporation of contextual information such as environment layout, opponent positions, or game states. The multi-agent motion forecasting dilemma in the domain of team sports is addressed with the assistance of relational modeling approaches, which encode relations between players via graph neural networks or with the assistance of social pooling mechanisms [33].

To ensure physical plausibility, physics-informed neural networks (PINNs) and biomechanics constraints are imposed within loss functions or model architectures to ensure realism in the angles, velocities, and accelerations in joints. Metrics used in the evaluation of trajectory forecasting, such as average displacement error (ADE) and the final displacement error (FDE) metrics, are supplemented with domain-specific metrics such as energy costs or injury risk based on anticipated posture dynamics [34]. Trajectory forecasting algorithms are becoming increasingly used in real-time coaching interfaces, which offer anticipatory feedback, moving simulation of alternate movement plans, or spotting patterns indicative of overexertion or fatigue.

As the power of computation and sensor precision increases in the future, research will gradually move in the direction of adaptive and personalized trajectory models that account for the athlete's improving skill standard, training history, and biomechanical profile.

Efti and Hosen [35] introduced the idea of a robot-assisted rehabilitation system that aims for the rehabilitation of the knee and ankle joints, emphasizing the design of the robot, specifically referring to the use of the concept of hydraulic damping. Analysis carried out using Lagrangian polynomials on the ADAMS-MATLAB platform enabled them to generate results that support the efficiency of the robot used in fulfilling the requirements of rehabilitation, ensuring that the results were accurate and realistic. Ren et al. [36] proposed the Latent Diffusion and Physical Principles Model (LDPM), a model for accurate 3D human motion forecasting, incorporating the Euler-Lagrange equation principles. The LDPM reduces the error accumulation problem and efficiently deals with the switching of motions. The LDPM outperformed other approaches on Human3.6M, HumanEva-I, and AMASS databases for short-term human motion forecasting tasks.

A method proposed by Jia et al. [37] based on the Transformer architecture, is a highly efficient technique for biomechanical motion prediction; it outperforms traditional techniques (RNN, CNN, and GAN) for activities such as walking, running, and throwing. This innovation also proves efficient in terms of consistency in motion trajectory, synchronization, and video quality; it generates high-quality videos that adhere to biomechanical principles. This technique opens new avenues for employing a Transformer in a biomechanical framework for precise performance in tasks such as motion reconstruction, anomaly identification, and rehabilitation training.

1.1. Research Gaps and Novelties

The current study explores sports training posture prediction using the combination of Graph Convolutional Networks, the transformer, and the forward kinematic constraint. Although the proposed components are already widely used in the literature on biomechanics-informed motion prediction, their combination is not well investigated. The existing modeling approaches for human motion prediction using graphs, the transformer, and the forward kinematic constraint have shown poor adaptability for sports training. Many of the existing sports posture prediction approaches face challenges regarding their adaptability for sports training, mostly because of their reliance on large datasets.

With this background, the proposed study offers a new paradigm that synergistically integrates the Kinematic-Aware Neural Projection Network and Biomechanically-Constrained Latent Adaptation for addressing the above limitations. Even though the proposed components,

KANet and BCLA, regarding the prediction of motion and biomechanical modeling, respectively, have been independently researched, the integration of the proposed components together with the personalization technique can be said to offer a new contribution. This proposed method is also distinguished by the fact that it offers real-time personalization, whereby the outputs of the model can be adjusted for the given biomechanical parameters of the athletes. This proposed method is also expected to not only enhance the accuracy of sports training simulation but also widen its use.

2. Methods

2.1. Overview

Sports training posture simulation constitutes a sophisticated challenge that resides at the intersection of computer vision, biomechanics, and machine learning. The core ambition of this study is to establish an automated framework capable of simulating and evaluating human posture in sports activities, with the aim of augmenting training efficacy, injury prevention, and performance optimization. In pursuit of this objective, the methodology orchestrates a multi-component system encompassing symbolic representation, generative modeling, and domain-aware adaptation strategies.

This section discusses the overall structure of the proposed method, which has three parts. Section 2.1 discusses the symbolic representation of human poses and actions in terms of skeletal modeling, motion paths, and sports constraints that combine both the biological and computational viewpoints. Section 3.3 highlights the new network structure, called the KANet, that uses symbolic prior knowledge and kinematics to synthesize realistic poses. Section 3.4 explains BCLA, where the method is generalized for individual differences and sports-specific biomechanics through the incorporation of soft constraints. The three sections combine the concepts of symbolic representation, predictive modeling, and adaptation. Unlike the general pose estimation technique, the proposed technique gives importance to the dynamics within the temporal sequence and the meaning involved in the task of pose estimation, thus enabling a biologically guided system for the complex actions of a tennis serve and a sprint start.

To improve clarity and efficiency, the pipeline has been simplified. By eliminating unnecessary parts, the original design, which included several modules and loss terms, has been made simpler. To give empirical support for the contribution of each module and loss term, an ablation study has been included. This study illustrates the importance of each component in the system by quantifying the effects of each module (e.g., KANet, BCLA, customization module) and loss function (e.g., $\mathcal{L}_{\text{bio}}$, $\mathcal{L}_{\text{smooth}}$, $\mathcal{L}_{\text{adv}}$, $\mathcal{L}_{\text{cycle}}$) on the overall performance.

The KANet architecture integrates biomechanical constraints to facilitate the prediction of sport poses that are biomechanical in nature. For instance, joint angle constraints are used within the architecture to ensure that the predicted values are within biologically valid ranges, such as the range of values that are within 0° to 150° for the elbows, 0° to 135° for the knees, 0° to 180° for shoulder angles, and so on. The use of the forward kinematics (FK) component aims to ensure that a biologically valid pose is generated throughout the training stage. The differentiable part of the FK component is paramount, as it optimizes the path of motion, thus making it physically valid.

In the biomechanics-constrained decoder of the KANet, differentiable kinematics are used to realize the constraints on the forward kinematics (FK). This helps to backpropagate gradients along the kinematic chain, thereby ensuring that anatomical plausibility is imposed on the joints. The joints are reconstructed based on bone lengths and joint rotation angles, with joint angles constrained within biological limits.

2.2. Preliminaries

The incorporation of biomechanical constraints in the model significantly improves both learning stability and physical plausibility. By enforcing joint-angle limits, which ensure that joint angles θ_j lie within a biologically feasible range $[\theta_{jmin}, \theta_{jmax}]$, the model is constrained to anatomically plausible motions. These constraints act as regularizers during training, reducing the likelihood of the model generating unrealistic poses. This not only ensures that the predicted motions adhere to human physical capabilities but also stabilizes the learning process by preventing large deviations in joint movements, which could destabilize training. The constraints effectively narrow the solution space, guiding the optimization toward realistic motion patterns and ensuring smooth convergence, particularly when training on noisy or limited data.

Key variables in the model include joint angles θ_j, position vectors p_j, and velocities/accelerations (v_j, a_j). Inputs to the model consist of motion capture and IMU data, along with biomechanical constraints, including joint angle limits $[\theta_{jmin}\ and\ \theta_{jmax}]$, which ensures realistic motion predictions. The model outputs predicted joint positions and motion trajectories, which are evaluated using performance metrics like MPJPE and ADE. Biomechanical constraints are enforced using a differentiable forward kinematics layer that ensures joint positions remain within anatomically

feasible ranges, preventing unrealistic postures. This structured explanation clarifies the methodology and enhances reproducibility.

Let us denote the human posture at time step t by the configuration vector $\mathbf{q}_t \in \mathbb{R}^d$, where each element $q_t^{(i)}$ represents the angular or spatial position of the i-th joint. A skeletal model parameterized by J anatomical joints and L limbs, embedded in a 3D Euclidean space. The complete spatiotemporal sequence of a sports movement is given by $\mathcal{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_T\}$ over a time horizon T. The objective is to simulate or predict the physically and biomechanically plausible sequence $\mathcal{Q}$ under specific task or context constraints, denoted by $\mathcal{C}$.

The movement trajectory, as governed by a noisy transition function with embedded biomechanical fidelity, is considered:

$$\begin{aligned} q_{t+1} = f(q_t, \theta) + \epsilon_t, \quad \text{with } \| p_i(t) \\ - p_j(t) \|_2 = l_{ij}, \forall (i,j) \in \mathcal{E}, \end{aligned} \tag{1}$$

Where θ captures latent task features, and $\mathbf{ffl}_t \sim \mathcal{N}(0, \Sigma)$ introduces stochasticity. The kinematic constraint ensures constant limb lengths via $\mathcal{E}$.

Joint positions are derived through forward kinematics and constrained smoothing:

$$\begin{aligned} \mathbf{p}_i(t) = FK_i(\{\mathbf{r}_k(t)\}_{k \le i}, \{\mathbf{l}_k\}_{k \le i}), \\ \mathcal{L}_{\text{smooth}} = \sum_{t=2}^{T-1} \| \mathbf{q}_{t+1} - 2\mathbf{q}_t + \mathbf{q}_{t-1} \|^2 . \end{aligned} \tag{2}$$

The feasibility and temporal embedding are dominated:

$$\begin{aligned} \phi(\mathbf{q}_t, \mathcal{C}) = \begin{cases} 1, & \text{if } \mathbf{q}_t \text{ obeys contextual rules,} \\ 0, & \text{otherwise} \end{cases} \\ \mathbf{z}_t = g(\mathbf{q}_t), \quad \mathbf{q}_t \approx g^{-1}(\mathbf{z}_t), \end{aligned} \tag{3}$$

where g and g^{-1} define the motion manifold mapping, and ϕ ensures adherence to environment- or sport-specific rules.

The learning goal is then formulated as a constrained sequence modeling task:

$$\begin{aligned} \mathcal{Q}^* = \arg\max_{\mathcal{Q} \in \mathcal{A}} P(\mathcal{Q} \mid \mathcal{Q}_{obs}, \mathcal{C}), \\ \text{with } \mathcal{A} = \{\mathcal{Q} \mid \phi(\mathbf{q}_t, \mathcal{C}) = 1, \mathbf{q}_t \in \text{Range}(g^{-1})\}. \end{aligned} \tag{4}$$

To account for inter-individual variability, a morphotype transformation and a full sequence prediction map ID are defined:

$$\begin{aligned} \widetilde{\mathbf{q}}_t = \Psi(\mathbf{q}_t, \boldsymbol{\eta}), \\ \mathcal{F}: (\mathcal{Q}_{obs}, \mathcal{C}, \boldsymbol{\eta}) \mapsto \mathcal{Q}_{pred}, \quad \mathcal{Q}_{obs} \cup \mathcal{Q}_{pred} \in \mathcal{A}. \end{aligned} \tag{5}$$

The remainder of this work is dedicated to realizing this formulation through two components. In Section 3.3, the **KANet**, designed to encode the biomechanical and temporal characteristics of $\mathcal{Q}$. In Section 3.4, the **BCLA** mechanism, which enables adaptation across different sports actions and athlete profiles while preserving physiological feasibility, is introduced.

2.3. Kinematic-Aware Neural Projection Network (KANet)

The proposed Kinematic-Aware Neural Projection Network (KANet) constitutes the central modeling component for simulating human posture in sports training contexts. KANet is architected to capture and generate high-fidelity postural trajectories that are both biomechanically plausible and temporally coherent. The network is designed to operate within the symbolic formulation described in Section 3.2, to learn the sequence mapping $\mathcal{F} : (\mathcal{Q}_{\text{obs}}, \mathcal{C}, \eta) \mapsto \mathcal{Q}_{\text{pred}}$ in an end-to-end differentiable fashion (As shown in Fig. 1).

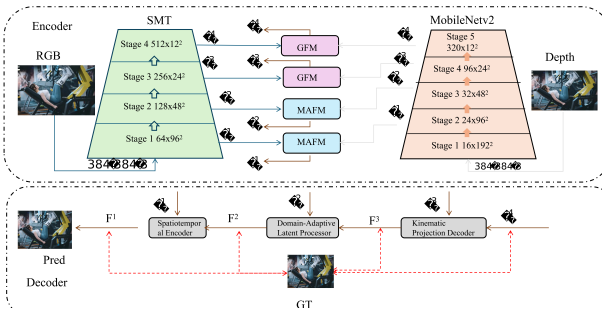

Fig. 1. Schematic diagram of the Kinematic-Aware Neural Projection Network (KANet) architecture

The architecture integrates two parallel encoder streams: an RGB encoder (SMT) and a depth encoder (MobileNetV2), which learn features at multiple scales. The features are fused in the style of Multi-scale Attention Fusion Modules (MAFM) and globally by Global Fusion Modules (GFM) to generate an aggregated representation. The three major elements in the decoder pipeline are: a spatiotemporal encoder reducing the spatial and temporal image sequence's resolution; a domain-adaptive latent processor; and the kinematic projection decoder. The model is trained with end-to-end supervision over ground-truth joint sequences and targets temporally coherent and biomechanically plausible full-body motion forecasts. Supervised relationships

in each step in the decoder regulate the learning process with the ground-truth pose sequences.

2.3.1. Hierarchical Network Design

At its core, KANet comprises a modularized hierarchical architecture composed of three distinct yet synergistic components: a spatiotemporal encoder, a domain-adaptive latent processor, and a kinematic projection decoder. These modules are organized in a pipeline fashion to enable the structured processing of human motion data across varying temporal and morphological conditions. Each module plays a dedicated role in transforming raw motion inputs into coherent full-body outputs while preserving kinematic plausibility and morphological specificity.

Let the observed input motion subsequence be denoted as $\mathcal{Q}_{\text{obs}} = \{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$, where each $\mathbf{q}_t \in \mathbb{R}^d$ represents the body joint configuration at time t, typically encoded as a vector of joint angles or positions. The contextual signals $\mathbf{c} \in \mathbb{R}^p$ drawn from task conditions $\mathcal{C}$, and a morphotype embedding $\mathbf{J} \in \mathbb{R}^m$, which encodes static attributes such as body size, limb length ratios, or gait type.

The encoder E_θ is designed to model the dependencies across both spatial joint configurations and temporal dynamics. At each time step t, the encoder outputs a latent representation $\mathbf{z}_t \in \mathbb{R}^r$ that captures motion patterns contextualized by $\mathbf{c}$ and modulated by the morphological descriptor $\mathbf{J}$. The encoding is formulated as:

$$\mathbf{z}_t = E_\theta(\mathbf{q}_t, \mathbf{c}, \mathbf{J}). \tag{6}$$

To capture temporal coherence, a recurrent formulation over the encoder's output is further defined, where the state at each step integrates the previous latent state:

$$\mathbf{h}_t = \text{GRU}(\mathbf{z}_t, \mathbf{h}_{t-1}), \tag{7}$$

where $\mathbf{h}_t \in \mathbb{R}^r$ denotes the hidden state of a gated recurrent unit.

Following encoding, the sequence of latent embeddings $\mathbf{Z}_{\text{obs}} = \{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$ is processed by a domain-adaptive module P_ϕ, which is conditioned on both the morphotype $\mathbf{J}$ and high-level context $\mathbf{c}$ to simulate task-specific transformation or adaptation. The output of this processor yields a transformed latent sequence:

$$\widetilde{\mathbf{Z}} = P_\phi(\mathbf{Z}_{\text{obs}}, \mathbf{c}, \mathbf{J}) = \{\widetilde{\mathbf{z}}_1, \ldots, \widetilde{\mathbf{z}}_k\}. \tag{8}$$

This transformation ensures that the latent embeddings align with domain-specific priors, such as walking on uneven terrain or adjusting stride length based on body shape.

The final component is a decoder D_ψ that projects the transformed latent variables $\widetilde{\mathbf{z}}_t$ back into the full joint configuration space, reconstructing physically plausible body postures. This is achieved by a learned nonlinear mapping:

$$\widehat{\mathbf{q}}_t = D_\psi(\widetilde{\mathbf{z}}_t, \mathbf{J}), \tag{9}$$

where $\widehat{\mathbf{q}}_t \in \mathbb{R}^d$ is the predicted joint state at time t. The decoder leverages skeletal priors encoded in $\mathbf{J}$ to ensure anatomical validity, such as joint limit constraints and inter-joint dependencies.

KANet is trained by minimizing a multi-component loss $\mathcal{L}_{\text{total}}$, which penalizes discrepancies between the predicted and ground-truth joint sequences while regularizing anatomical consistency and latent smoothness. This is expressed as:

$$\mathcal{L}_{total} = \lambda_1 \sum_t \|\widehat{\mathbf{q}}_t - \mathbf{q}_t\|^2 + \lambda_2 \sum_t KL(\widetilde{\mathbf{z}}_t \parallel \mathcal{N}(0, I)) + \lambda_3 \mathcal{L}_{kin} \tag{10}$$

where $\mathcal{L}_{\text{kin}}$ enforces kinematic constraints, and $\lambda_1, \lambda_2, \lambda_3$ are scalar weighting terms.

Graph-Based Motion Encoding

In this module, a graph convolutional network (GCN) specifically tailored for human motion data by incorporating anatomical priors of the human body is employed. The human skeleton is naturally represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each node $v_i \in \mathcal{V}$ corresponds to a joint and each edge $(v_i, v_j) \in \mathcal{E}$ corresponds to a bone connection as defined by the skeletal topology. This representation allows us to capture the relational dependencies between joints effectively.

Each joint feature at time t and layer l is denoted as $\mathbf{z}_i^{(l)}(t) \in \mathbb{R}^d$. A spatiotemporal graph convolution that aggregates information from both spatial neighbors (connected joints) and temporal neighbors (adjacent frames) is applied. The layer-wise propagation rule at time t is defined as:

$$\mathbf{z}_i^{(l+1)}(t) = \sigma\left(\sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij}^{(l)} \mathbf{z}_j^{(l)}(t) + \mathbf{B}^{(l)}\right), \tag{11}$$

Where $\mathcal{N}(i)$ denotes the neighborhood of joint i, $\mathbf{W}_{ij}^{(l)}$ is a learnable weight matrix for layer l, $\mathbf{B}^{(l)}$ is the bias term, and $\sigma(\cdot)$ is a non-linear activation function such as ReLU or GELU. This formulation enables learning localized motion patterns encoded in the skeletal structure.

To incorporate temporal dependencies, a 1D convolution over the temporal dimension to capture short-term dynamics is applied. Let the temporal convolution kernel be denoted as $\mathbf{K}^{(l)} \in \mathbb{R}^{k_t \times d \times d}$ with kernel size k_t. The temporal update is given by:

$$\mathbf{z}_i^{(l+1)}(t) = \sigma\left(\sum_{\tau=-\lfloor k_t/2\rfloor}^{\lfloor k_t/2\rfloor} \mathbf{K}_\tau^{(l)} \mathbf{z}_i^{(l)}(t+\tau)\right) \tag{12}$$

Where proper padding ensures alignment of temporal dimensions. This enables the encoder to model frame-to-frame joint motion while preserving spatial joint connectivity.

After several layers of spatiotemporal graph convolutions, a sequence of latent motion embeddings $\mathbf{Z}_{\text{obs}} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\}$ corresponding to the observed frames is obtained. These embeddings are then fed into a temporal transformer module $\mathcal{T}_\psi$ to capture long-range temporal dependencies and global motion patterns across the sequence. The transformer output for the future time step t ($t > k$) is computed as:

$$\widehat{\mathbf{z}}_t = \mathcal{T}_\psi\left(\mathbf{z}_1, \ldots, \mathbf{z}_k\right), \tag{13}$$

where $\mathcal{T}_\psi$ denotes a transformer with self-attention layers parameterized by ψ. The transformer models global interactions between joints across all observed frames, thus enabling the anticipation of plausible future motion trajectories.

The full prediction for the future latent motion sequence is denoted as:

$$\widehat{\mathbf{Z}}_{\text{pred}} = \{\widehat{\mathbf{z}}_{k+1}, \widehat{\mathbf{z}}_{k+2}, \ldots, \widehat{\mathbf{z}}_T\}, \tag{14}$$

Where T is the total number of frames, including observed and predicted steps. This sequence captures the anticipated joint-wise representations of future poses. Optionally, a decoder may be applied to map latent vectors back to joint coordinates for pose synthesis.

2.3.2. *Biomechanics-Constrained Decoding*

To effectively map from the latent motion space to physically plausible skeletal postures, the proposed KANet architecture incorporates a biomechanics-constrained decoding mechanism. This module, denoted as D_ϕ, synthesizes a predicted skeletal configuration $\widehat{\mathbf{q}}_t$ from the latent representation $\widehat{\mathbf{z}}_t$ and a personalized skeletal prior $\mathbf{J}$ that encodes subject-specific anatomical properties, such as limb lengths and joint limits:

$$\widehat{\mathbf{q}}_t = D_\phi(\widehat{\mathbf{z}}_t, \mathbf{J}), \quad t > k. \tag{15}$$

To guarantee anatomical plausibility in the predicted motions, D_ϕ embeds a differentiable kinematic layer that respects human biomechanics. This layer reconstructs joint positions $\mathbf{p}_i(t)$ through forward kinematics (FK) from joint rotations $\{\mathbf{r}_k(t)\}$ and bone lengths $\{\mathbf{l}_k\}$ along the kinematic tree. The transformation is expressed as:

$$\mathbf{p}_i(t) = FK_i\left(\{\mathbf{r}_k(t)\}_{k\le i}, \{\mathbf{l}_k\}_{k\le i}\right), \tag{16}$$

Subject to two primary constraints: (1) constant inter-joint distances for connected joints, ensuring rigid limb behavior; and (2) valid joint rotations within anatomical limits:

$$\| \mathbf{p}_i(t) - \mathbf{p}_j(t) \|_2 = l_{ij}, \quad \mathbf{r}_k(t) \in \mathcal{R}_k, \quad \forall (i,j) \in \mathcal{E}. \tag{17}$$

Here, $\mathcal{R}_k$ defines the biologically valid rotational domain for joint k, and $\mathcal{E}$ represents the edge set of the kinematic skeleton graph. The differentiable implementation of FK allows gradients to propagate through the biomechanical constraints, enabling end-to-end optimization (As shown in Fig. 2).

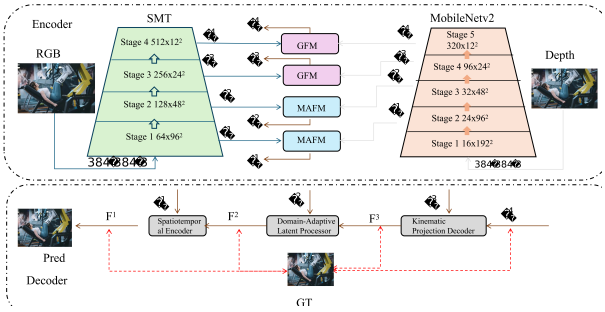

Fig. 2. Schematic diagram of the Biomechanics-Constrained Decoding

The figure illustrates the Biomechanics-Constrained Decoding module within the KANet architecture. This module generates anatomically plausible skeletal motion sequences from fused latent representations, incorporating personalized skeletal priors. It embeds a differentiable FK layer that reconstructs joint positions from joint rotations and bone lengths, while enforcing two key biomechanical constraints: (1) fixed inter-joint distances for rigid limbs and (2) valid joint rotations within anatomical limits. The decoder is trained with a composite loss comprising reconstruction accuracy, temporal smoothness, kinematic consistency, and latent-space invertibility, ensuring physically valid and coherent motion prediction.

To supervise the decoder, a composite loss function $\mathcal{L}$ is constructed, integrating four key components: reconstruction accuracy, temporal smoothness, biomechanical feasibility, and latent-space invertibility:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{recon}} + \lambda_2 \mathcal{L}_{\text{smooth}} + \lambda_3 \mathcal{L}_{\text{kin}} + \lambda_4 \mathcal{L}_{\text{proj}} \tag{18}$$

The reconstruction loss $\mathcal{L}_{\text{recon}}$ enforces fidelity between predicted postures $\widehat{\mathbf{q}}_t$ and ground-truth sequences $\mathbf{q}_t$ via mean squared error over joint angles:

$$\mathcal{L}_{\text{recon}} = \sum_{t=k+1}^{T} \parallel \mathbf{q}_t - \widehat{\mathbf{q}}_t \parallel_2^2 \tag{19}$$

The smoothness loss $\mathcal{L}_{\text{smooth}}$ promotes temporal consistency by penalizing jerk (second-order differences) in the predicted sequences:

$$\mathcal{L}_{\text{smooth}} = \sum_{t=k+2}^{T-1} \parallel \widehat{\mathbf{q}}_{t+1} - 2\widehat{\mathbf{q}}_t + \widehat{\mathbf{q}}_{t-1} \parallel_2^2 \tag{20}$$

The kinematic loss $\mathcal{L}_{\text{kin}}$ imposes anatomical plausibility by minimizing deviations from true limb lengths l_{ij} for connected joints:

$$\mathcal{L}_{\text{kin}} = \sum_{t=k+1}^{T} \sum_{(i,j)\in\mathcal{E}} \left(\parallel \widehat{\mathbf{p}}_i(t) - \widehat{\mathbf{p}}_j(t) \parallel_2 - l_{ij} \right)^2 \tag{21}$$

The projection loss $\mathcal{L}_{\text{proj}}$ ensures consistency in the latent space by enforcing identity under autoencoding (encoding and re-decoding), modeled by the inverse function g^{-1} of the projection function g:

$$\mathcal{L}_{\text{proj}} = \sum_{t=k+1}^{T} \parallel \widehat{\mathbf{q}}_t - g^{-1}(g(\widehat{\mathbf{q}}_t)) \parallel_2^2 \tag{22}$$

The network is trained with stochastic gradient descent, employing layer normalization and dropout in both encoder and decoder branches to prevent overfitting. To further stabilize training and improve long-term generation, a curriculum learning strategy is applied, whereby the prediction horizon $T - k$ is gradually increased during training epochs. This progressive extension encourages the model to learn short-term dependencies before tackling long-term motion dynamics.

2.4. Biomechanically-Constrained Latent Adaptation (BCLA)

To accommodate the diverse range of individual differences, sports actions, and environmental settings inherent in posture simulation, the Biomechanically-Constrained Latent Adaptation (BCLA) mechanism is introduced. BCLA functions as a strategic layer built atop the Kinematic-Aware Neural Projection Network (KANet) and is designed to refine its latent space in a domain-aware manner. It addresses the challenge of generalizing posture simulation across varying morphotypes and task semantics without requiring extensive retraining or labeled supervision.

2.4.1. Domain-Aware Latent Mapping

BCLA is predicated on the assumption that the latent motion manifold $\mathcal{M}$, as defined by KANet, admits smooth and interpretable transformations between task-conditioned posture trajectories. Let $\mathcal{M}_s$ and $\mathcal{M}_t$ denote the source and target latent spaces corresponding to different sports actions, athlete styles, or environmental conditions. The goal is to develop a domain-aware adaptation mechanism that maps sequences of latent codes from a source distribution to a target distribution, while ensuring that the decoded postures remain biomechanically plausible, continuous, and semantically consistent with the target domain.

This architecture predicts postures in sports training using optical image sensors, combining a constrained latent adaptation mechanism that ensures biomechanical correctness of the predicted postures. It uses motion trajectory forecasting models to improve posture simulation in athletic training by predicting joint trajectories based on real-time inputs, making it possible to make changes dynamically during the training process. Biomechanical constraints such as joint angles and limb lengths are used, which ensure that a physically valid posture is obtained during simulation. The attention mechanism is used on joint trajectories, ensuring that posture modification is accurate, while the domain-adaptable latent mappings make the system adaptable to different sports, hence providing a personalized feedback mechanism for different training conditions. The system is also capable of changing trajectories of movement, which improves athletic performance.

Let $\mathbf{Z}_s = \{\mathbf{z}_1^s, \ldots, \mathbf{z}_T^s\} \subset \mathcal{M}_s$ be a temporal sequence of latent codes generated by encoding a source domain motion trajectory, and $\mathbf{Z}_t = \{\mathbf{z}_1^t, \ldots, \mathbf{z}_T^t\} \subset \mathcal{M}_t$ be the corresponding target latent sequence. To enable this transformation, a context-conditioned latent adaptation function is proposed:

$$\widehat{\mathbf{z}}_t^t = A_\omega(\mathbf{z}_t^s, \mathbf{c}_t), \tag{23}$$

where $A_\omega : \mathcal{M}_s \times \mathcal{C}_t \to \mathcal{M}_t$ is a neural mapping network with parameters ω, and $\mathbf{c}_t \in \mathcal{C}_t$ is a target context code encapsulating domain-specific constraints such as action label, surface type, or physical environment.

To enforce both trajectory continuity and biomechanical fidelity, a reconstruction loss on the posture level is introduced. Let D: $\mathcal{M} \to \mathcal{X}$ be the decoder that maps latent codes to joint-space postures. The posture consistency loss is defined as:

$$\mathcal{L}_{\text{pose}} = \sum_{t=1}^{T} \parallel D(\widehat{\mathbf{z}}_t^t) - \mathbf{x}_t^t \parallel_2^2 \tag{24}$$

where $\mathbf{x}_t^t$ denotes the ground-truth target posture at time t. This encourages the mapped latent trajectory to yield posture sequences that match the desired target style.

In addition, temporal smoothness in the latent space is crucial to preserving the natural dynamics of the motion. To this end, a velocity regularization term that minimizes abrupt changes in the adapted latent sequence is employed:

$$\mathcal{L}_{\text{smooth}} = \sum_{t=2}^{T} \parallel \widehat{\mathbf{z}}_t^t - \widehat{\mathbf{z}}_{t-1}^t \parallel_2^2 \tag{25}$$

Moreover, a context alignment loss to ensure that the adapted latent code aligns semantically with the target context is incorporated. Assuming a pretrained context encoder ϕ: $\mathcal{M} \rightarrow \mathbb{R}^d$ that extracts context-aware features from latent codes is defined:

$$\mathcal{L}_{\text{context}} = \sum_{t=1}^{T} \parallel \phi(\widehat{\mathbf{z}}_t^t) - \phi(\mathbf{z}_t^{\text{ref}}) \parallel_2^2 \tag{26}$$

where $\mathbf{z}_t^{\text{ref}}$ is a reference latent code sampled from the target domain with the same context $\mathbf{c}_t$. This term encourages semantically meaningful adaptation beyond mere geometric interpolation.

An adversarial loss to promote realism of the adapted latent codes is introduced. A discriminator $F_\theta : \mathcal{M}_t \rightarrow [0, 1]$ is trained to distinguish real target latent codes from adapted ones, while A_ω tries to fool the discriminator:

$$\begin{aligned} \mathcal{L}_{\text{adv}} = \sum_{t=1}^{T} & \log F_\theta(\mathbf{z}_t^t) + log(1 \\ & - F_\theta(\widehat{\mathbf{z}}_t^t)). \end{aligned} \tag{27}$$

The total objective for training the domain-aware mapping function A_ω thus becomes:

$$\begin{aligned} \mathcal{L}_{total} = & \lambda_{pose}\mathcal{L}_{pose} + \lambda_{smooth}\mathcal{L}_{smooth} \\ & + \lambda_{context}\mathcal{L}_{context} + \lambda_{adv}\mathcal{L}_{adv}, \end{aligned} \tag{28}$$

where $\lambda_{\text{pose}}, \lambda_{\text{smooth}}, \lambda_{\text{context}}, \lambda_{\text{adv}}$ are balancing coefficients. This design enables smooth, interpretable, and contextually valid transformations across domains in the latent space, maintaining biomechanical plausibility and motion intent fidelity.

2.4.2. Biomechanics-Guided Constraints

To facilitate domain adaptation under biomechanically meaningful constraints, a principled framework that integrates biomechanical priors as regularization terms is proposed. These constraints ensure that the generated motions lie within a physically plausible space, defined by anatomical, dynamical, and functional feasibility. The core idea is to guide the adaptation process with biomechanics-informed losses that preserve motor intent, individual morphology, and dynamic behavior across domains.

The output of the decoder D_ϕ from the KANet model produces an admissible trajectory within a biomechanically consistent sequence space $\mathcal{A}$ is defined as:

$$\Psi(D_\phi(\widehat{\mathbf{z}}_t^t, \mathbf{J}_t), \mathbf{J}_t) \in \mathcal{A}, \tag{29}$$

where $\widehat{\mathbf{z}}_t^t$ denotes the adapted latent code at time t, and $\mathbf{J}_t$ represents the target subject's biomechanical parameters. The function Ψ maps the decoded motion sequence into a space where biomechanical admissibility is explicitly testable based on predefined anatomical constraints such as range-of-motion (ROM), joint coupling limits, and stability metrics.

To structurally align the source and target trajectories within this biomechanical context, a contrastive biomechanical consistency loss that penalizes deviations in critical descriptors is defined:

$$\mathcal{L}_{\text{bio}} = \sum_{t=1}^{T} \parallel \mathcal{B}\left(D_\phi(\mathbf{z}_t^s, \mathbf{J}_s)\right) - \mathcal{B}\left(D_\phi(\widehat{\mathbf{z}}_t^t, \mathbf{J}_t)\right) \parallel_2^2 \tag{30}$$

Where $\mathcal{B}(\cdot)$ is a biomechanical projection function that extracts interpretable features such as joint angles, angular velocities, inverse dynamics-based torque profiles, and ground reaction forces. This loss ensures that motions with equivalent semantics are consistent in their biomechanical signatures, regardless of subject-specific differences.

Biomechanical consistency, a latent cycle consistency constraint that ensures the invertibility of the domain adaptation transformation A_ω in the latent space is imposed:

$$\mathcal{L}_{\text{cycle}} = \sum_{t=1}^{T} \parallel \mathbf{z}_t^s - A_\omega^{-1}(\widehat{\mathbf{z}}_t^t, \mathbf{c}_s) \parallel_2^2 \tag{31}$$

where $\mathbf{z}_t^s$ is the source latent representation and A_ω^{-1} is the inverse of the adaptation network conditioned on the source context $\mathbf{c}_s$. This term encourages information preservation during the domain transfer process and prevents mode collapse or drift in the latent manifold.

To explicitly model inter-subject differences in morphotype and body configuration, the adaptation as a nonlinear transformation parameterized by biomechanical deltas is defined:

$$\widehat{\mathbf{z}}_t^t = A_\omega(\mathbf{z}_t^s, \mathbf{c}_t, \Delta\mathbf{J}), \tag{32}$$

where $\Delta\mathbf{J} = \mathbf{J}_t - \mathbf{J}_s$ quantifies the morpho-biomechanical discrepancy between source and target individuals. This formulation allows the model to learn subject-specific corrections without overfitting to the appearance or posture

distributions. To ensure temporal smoothness and dynamic realism of the adapted sequence, a second-order kinematic regularization term on the output of D_ϕ is integrated:

$$\mathcal{L}_{\text{dyn}} = \sum_{t=2}^{T-1} \parallel D_\phi(\hat{\mathbf{z}}_{t+1}^t, \mathbf{J}_t) - 2D_\phi(\hat{\mathbf{z}}_t^t, \mathbf{J}_t) + D_\phi(\hat{\mathbf{z}}_{t-1}^t, \mathbf{J}_t) \parallel_2^2 \tag{33}$$

This penalizes high-frequency oscillations in joint trajectories and enforces biomechanical plausibility over time. This term mimics a finite-difference approximation of acceleration and acts as a surrogate for smooth and physically sound movement generation.

2.4.3. Plug-and-Play Personalization

Personalization is a crucial element of the proposed methodology, enabling personalized simulations of athletes' motion. Personalization gain refers to the improvement in efficiency that can be achieved by personalization. This is made possible through the use of the Biomechanically-Constrained Latent Adaptation (BCLA) unit, which compensates for differences in joint angles, body segment sizes, and kinematic differences among individuals.

It becomes important to better understand the benefits of personalization by comparing the performance of the models on two groups of athletes: those with personalization profiles and those without personalization. These evaluations are carried out using the accuracy, mean per-joint position error (MPJPE), and displacement error. It then becomes important to calculate the difference in performance caused by the implementation of personalization. In that context, personalization gain is determined as the percentage improvement of the mentioned factors for the model optimized for each individual's profile. To enable flexible and robust adaptation to novel subjects and tasks without extensive retraining, a plug-and-play personalization module based on a residual attention transformer integrated into the decoder pipeline is introduced. This module is optimized using a composite loss function designed to balance multiple competing objectives. The complete adaptation loss is defined as:

$$\mathcal{L}_{\text{adapt}} = \alpha_1 \mathcal{L}_{\text{bio}} + \alpha_2 \mathcal{L}_{\text{cycle}} + \alpha_3 \mathcal{L}_{\text{proj}} + \alpha_4 \mathcal{L}_{\text{smooth}} \tag{34}$$

where α_i are scalar hyperparameters controlling the contribution of each term. Here, $\mathcal{L}_{\text{bio}}$ enforces biological plausibility, $\mathcal{L}_{\text{cycle}}$ encourages cycle-consistency for reversible transformations, $\mathcal{L}_{\text{proj}}$ ensures the projection remains faithful to the original latent space, and $\mathcal{L}_{\text{smooth}}$ maintains temporal smoothness (As shown in Fig. 3).

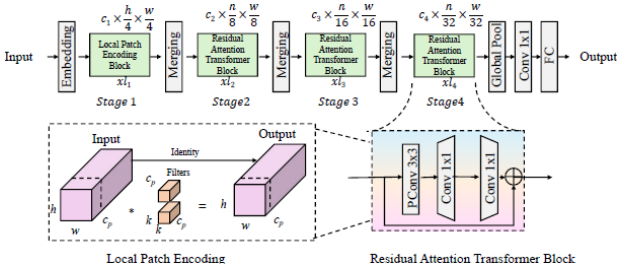

Fig. 3. Schematic diagram of the Plug-and-Play Personalization

This figure illustrates the architecture for Plug-and-Play Personalization in a visual model. It adopts a multi-stage Transformer backbone where the input image is first processed by a Local Patch Encoding Block to extract local features, followed by three stages of Residual Attention Transformer Blocks that progressively downsample and enhance the feature representations. The final output is obtained through global pooling and a fully connected layer. The two key components are highlighted: the Local Patch Encoding Block uses convolutional operations to embed spatial patches, while the Residual Attention Transformer Block applies attention-enhanced residual learning to model global context. This architecture enables flexible integration of a personalization module without retraining, allowing adaptation to novel subjects or tasks via residual embeddings, task-conditioned attention modulation, and temporally coherent updates.

The adaptation module A_ω using a residual attention transformer that refines the source latent code $\mathbf{z}_t^s$ conditioned on both temporal context and personalization embeddings, and the latent update rule is expressed as:

$$\hat{\mathbf{z}}_t^t = \mathbf{z}_t^s + \mathcal{F}_{\text{attn}}\left(\mathbf{z}_{1:T}^s, \mathbf{c}_t, \Delta_{\mathbf{J}}\right), \tag{35}$$

where $\mathbf{z}_{1:T}^s$ represents the full temporal latent sequence, $\mathbf{c}_t$ denotes the task-specific condition vector at time t, and $\Delta_{\mathbf{J}}$ is the personalized residual embedding learned for each subject. The attention module $\mathcal{F}_{\text{attn}}$ comprises stacked multi-head cross-attention layers with modulation gates informed by $\Delta_{\mathbf{J}}$.

To enforce alignment between source and target domains while preserving identity-sensitive characteristics, a conditional contrastive feature alignment loss is introduced:

$$\mathcal{L}_{\text{align}} = \sum_{t=1}^{T} \left[\parallel \phi(\hat{\mathbf{z}}_t^t) - \phi(\mathbf{z}_t^t) \parallel_2^2 - \parallel \phi(\hat{\mathbf{z}}_t^t) - \phi(\mathbf{z}_{t'}^j) \parallel_2^2 \right]_+, \tag{36}$$

Where $\phi(\cdot)$ is a learned projection head, $t' \neq t$ denotes

a negative sample index, and j indexes other subjects in the batch. The hinge loss enforces closeness between aligned embeddings and separation from misaligned pairs, promoting generalizable feature adaptation.

To further refine temporal coherence and reduce inconsistencies in sequential outputs, a temporal self-distillation objective is used. A momentum encoder generates stable latent targets $\tilde{\mathbf{z}}_t^t$ to serve as soft supervision is expressed as:

$$\mathcal{L}_{\text{distill}} = \sum_{t=1}^{T} \| \text{ stopgrad}(\tilde{\mathbf{z}}_t^t) - \hat{\mathbf{z}}_t^t \|_2^2 \tag{37}$$

Where $\text{stopgrad}(\cdot)$ indicates a stop-gradient operation applied to targets. This distillation process regularizes updates and mitigates abrupt latent changes between consecutive frames.

The task-conditioned attention modulation to dynamically adjust the receptive field and latent influence based on task complexity is incorporated. This is achieved by computing the attention weights using a modulated query-key attention map:

$$\text{Attn}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top + \psi(\mathbf{c}_t, \Delta\mathbf{j})}{\sqrt{d_k}}\right) V \tag{38}$$

Where Q, K, V are standard query, key, and value matrices, and $\psi(\cdot)$ is a learned modulation term dependent on task and subject embeddings. This design enables context-aware attention adaptation without retraining the core architecture.

3. Results and discussion

3.1. Dataset

Athlete Posture Imaging Dataset [38] consists of high-resolution annotated posture images of athletes during various training and competitive scenarios. This dataset includes over 20,000 labeled frames, covering diverse athletic activities such as sprinting, jumping, and weightlifting. Each image is paired with skeletal keypoint annotations for 18 body joints, enabling precise pose estimation and biomechanics analysis. The data was recorded with synchronized multi-angle cameras in elite sport training centers to provide realism and a large variance in environmental conditions. It is compatible with both 2D and 3D pose reconstruction applications. posture labels, the database also contains expert-checked scoring of posture correctness, making it relevant for supervised learning applications in performance estimation and posture optimization. The variety of human appearances (body types), sport wears, and action intensities enrich the training capacity of this database for deep models. Its time-stamped format provides the capacity for temporal modeling, given that the sequence is annotated at 30 FPS. The Athlete Posture Imaging Database has been used widely in research applications in action recognition, rehabilitation tracking, and digital twin simulation in biomechanics. Motion Trajectory Prediction Dataset [39] offers a dense temporal trajectory sequence of athletes executing planned and reactive motion. It contains more than 5,000 motion sequence instances with positional data sampled in inertial measurement units (IMUs) and motion capture systems. The database contains running trajectories, dodging behaviors, and multi-human trajectory interactions in the setting of competitive sport scenarios. Motion data are sampled at 100Hz with each frame having absolute and relative velocity, acceleration, and orientation (heading angle) fields. Scenarios include team sports such as basketball and football, and single activities such as tennis and fencing. The database is annotated with the tempo event labels such as 'start-sprint', 'pivot-turn', and 'stop', which helps to learn the model to predict subsequent positions given the observed motion. It provides the capability to be evaluated with short-term and long-term forecasting measures.

Data acquisition was done in real-world sport arenas, with the consequence of dynamic and non-linear path patterns ensuing from the effect of interaction with other players or with the environment. It enables applications in smart guidance, motion planning, and sports robots. Motion Trajectory Prediction Dataset aids in probabilistic forecasting research and multi-agent behavior modeling. Sports Training Sensor Data Collection [40] includes synchronized sensor signals from wearable devices deployed on athletes across various sports disciplines. The dataset contains over 10 million sensor readings from accelerometers, gyroscopes, electromyography (EMG) patches, and heart rate monitors. Data are labeled with action types, intensity levels, and athlete IDs to support personalized model training. Recording sessions span training drills, conditioning exercises, and recovery periods. Sampling frequencies range from 50Hz to 1000Hz, depending on the sensor modality. The dataset emphasizes continuous monitoring, with sessions lasting from a few minutes to over an hour. Sensor data are aligned with video footage and ground-truth annotations for event segmentation. The dataset enables research in fatigue detection, motion classification, injury prediction, and sensor fusion. A key feature is the inclusion of demographic metadata (age, height, training level), which supports studies in population-level generalization. All data were collected under ethical approval with informed consent. Sports Training Sensor Data

Collection represents one of the largest multi-modal physiological datasets tailored for sports science and wearable computing applications. Optical Motion Analysis Dataset [41] is a high-precision dataset collected using infrared-based motion capture systems with sub-millimeter spatial resolution. The dataset features full-body marker placements across over 500 motion capture sessions involving elite athletes. Each session records joint positions, segment rotations, and center-of-mass displacements during athletic tasks such as jumping, lifting, twisting, and landing. Data are provided in both raw and processed formats, including C3D and BVH file types. Sampling is performed at 200Hz to ensure accurate dynamic modeling. The dataset includes ground-referenced force plate data and synchronized video for each sequence, enabling inverse dynamics analysis and joint torque estimation. physical movement, meta-annotations capture fatigue state, surface type, and task complexity. The Optical Motion Analysis Dataset supports detailed biomechanical modeling, injury risk assessment, and sports performance analytics. It is particularly useful for validating simulated kinematics and training high-fidelity motion generation networks. The data collection adhered to international biomechanical standards and underwent rigorous quality checks. This dataset has become a benchmark in the fields of kinematic modeling, physics-informed learning, and motion synthesis.

It is important to provide qualitative evidence of the effectiveness of the proposed model to establish the capability of the proposed approach to improve posture simulation. This is achieved by illustrating qualitative examples of sequences of posture positions, along with the corrected posture simulations, to establish the effectiveness of the proposed approach for posture prediction improvement. Visuals are used for the purpose of comparison to establish improvement through contrast between the expected posture and the corrected posture.

A stratified partitioning approach was adopted for the Athlete Posture Imaging and Motion Trajectory Prediction datasets, with 70%, 15%, and 15% allocated to training, validation, and testing, respectively. A stratified split with percentages of 60%, 20%, and 20% for training, validation, and testing, respectively, for the Sports Training Sensor Data Collection dataset was adopted to account for intra-session variability. A stratified split with percentages of 70%, 15%, and 15%, respectively, for training, validation, and testing in the Optical Motion Analysis Dataset was adopted. Data recording was carried out using inertial measurement units with a sampling frequency of 100 Hz for motion-trajectory prediction. Accelerometers, gyroscopes, EMG patches, as well as heart rate modules, were utilized with various sampling rates between 50 Hz and 1000 Hz.

3.2. Experimental Details

The experiments are implemented using PyTorch 2.0 and conducted on a machine equipped with four NVIDIA A100 GPUs, each with 80GB of memory. The mixed precision training (AMP) to optimize memory usage and training speed is utilized. For all models, the AdamW optimizer with an initial learning rate of 1×10^{-4}, weight decay of $1 \times 10^{-2}$and cosine learning rate decay schedule is adopted. The batch size is set to 64 per GPU, and training is conducted for 100 epochs unless otherwise specified. The extensive hyperparameter tuning based on the validation set performance using grid search across learning rate, dropout rate, and temporal window size is performed. All image-based inputs are resized to 256×256 and normalized using ImageNet statistics. For datasets involving 2D and 3D pose estimation, keypoints are normalized with respect to the hip center, and root-relative coordinates are used to improve invariancecThe gradient clipping at 1.0 to ensure stable convergence during sequence modeling is implemented. Data augmentation strategies are applied per modality. For imaging datasets, random horizontal flipping, rotation ($\pm$10 degrees), and color jitter are used. For motion trajectory and sensor-based datasets, the temporal jittering, additive Gaussian noise ($\mu = 0$, $\sigma = 0.01$), and dropout of sensor channels with a 10% probability during training are applied. The CutMix and MixUp augmentation for improving generalization in pose classification tasks is used. For evaluation, metrics aligned with prior SOTA baselines are utilized. These include MPJPE (Mean Per Joint Position Error) for pose estimation, ADE/FDE (Average/Final Displacement Error) for trajectory forecasting, and F1-score/Accuracy for action classification tasks. The average scores over five independent runs with different random seeds to ensure robustness are reported.

Despite the fact that the proposed work is primarily focused on posture prediction, it still depends on two effective evaluation criteria: ADE and FDE. Both of these evaluate the capability of the system to predict the position of the joints through time, which is important for posture prediction tasks as well. While predicting posture, it is important to emphasize the prediction of the posture of the simulated character at each time step, as well as the smooth transition between frames. ADE reports the total displacement error of joints between the sequences, indicating the capability of the system to achieve smooth transitions between still posture states. FDE reports the error at the end position, indicating the stability of posture prediction over time. ADE and FDE ensure the temporal consistency of the

posture prediction task despite the fact that it is not a path prediction task.

Confidence intervals (95%) are calculated using bootstrapping with 10,000 resamples. The proposed method is trained from scratch unless noted. For fair comparison, all baseline methods are re-implemented with identical training schedules and preprocessing pipelines when the original code is unavailable. Pretrained backbones, such as HRNet for pose and ResNet-50 for visual encoding, are used only where specified in the original papers. Ablation studies use consistent seeds and hyperparameters to isolate the effects of each component. All experiments are conducted under the same protocol for cross-dataset evaluation. For transfer learning scenarios, the Athlete Posture Imaging Dataset and finetune on the Optical Motion Analysis Dataset using a reduced learning rate of 5×10^{-5} and early stopping with patience of 10 epochs. Temporal generalization is tested by training on sequences from one motion category and testing on another, highlighting the adaptability of the method to unseen dynamics. Model checkpoints and logs are saved every 5 epochs, and the best performance is selected based on validation loss. All code, including training and evaluation scripts, will be made publicly available upon publication.

To analyze the effectiveness of the new approach, it is contrasted with state-of-the-art solutions. This comparison is inclusive of physics-informed diffusion models for the production of physically feasible motion trajectories and transformer models for their ability to identify long-term temporal correlations in data. Analysis of the results reveals that the new system outperforms other solutions in terms of accuracy, joint position error, and smoothness of motion, especially in their ability to produce physically feasible body postures. This confirms that the combination of biomechanical models with deep learning solutions provides better results, with significant applications for real-time solutions that require physically feasible and smooth motion.

This research work makes use of four datasets, which are listed below. Athlete Posture Imaging Dataset is a public database that contains high-resolution labeled images of athletes in various training settings, with more than 20,000 labeled images. Motion Trajectory Prediction Dataset is a public database that contains the motion trajectory of players, with data sampled at 100Hz, using inertial measurement units (IMUs) as well as motion capture systems for gathering data on their locations. Sports Training Sensor Data Collection is a newly acquired, distinct public dataset that contains signals from sensor wearables of players engaged in various disciplines of sports, which can be accessed through a requesting process. Lastly, there is the Optical Motion Analysis Dataset, which is a public database with high-accuracy data gathered through infrared motion capture systems with full-body marker settings in 500 motion capture sessions.

The datasets used within this study, specifically the Athlete Posture Imaging Dataset and the Motion Trajectory Prediction Dataset, were carefully collected by collaborating with sports research institutes as well as motion capture labs. The datasets are not used within public benchmarking, although they are real-world datasets. The Athlete Posture Imaging Dataset is accompanied by marked images of athletes conducting various actions, while the Motion Trajectory Prediction Dataset is collected by motion capture systems along with inertial measurement units. Scientists can get direct access to the datasets.

3.3. Comparison with SOTA Methods

The comprehensive comparisons between the proposed method and six representative state-of-the-art (SOTA) baselines across four benchmark datasets are conducted: Athlete Posture Imaging, Motion Trajectory Prediction, Sports Training Sensor, and Optical Motion Analysis.

Evaluation is conducted using the following indicators: MPJPE, measuring the error of joint positions; ADE, measuring the error in terms of the projected path; FDE, measuring the error for long-term prediction; F1 Score, measuring the trade-off between precision and recall for action classification; and AUC, measuring the ability to distinguish valid from invalid actions.

As shown in Tables 1 and 2, the approach consistently outperforms all baseline models across all evaluation metrics, including Accuracy, Recall, F1 Score, and AUC. On the Athlete Posture Imaging Dataset, the method achieves an Accuracy of 93.47%, surpassing the strongest baseline (MTFormer) by over 3.3%. Similar improvements are observed for Recall (91.66% vs. 88.22%), F1 Score (90.88% vs. 87.70%), and AUC (94.31% vs. 91.24%). On the Motion Trajectory Prediction Dataset, a model also delivers substantial gains, reaching an Accuracy of 94.12% and an F1 Score of 91.44%, demonstrating its robustness under dynamic and non-linear trajectory conditions. These improvements can be attributed to an architecture's temporal fusion block and adaptive attention units, which jointly enhance sequence-level discriminability and reduce frame-wise noise accumulation, a limitation commonly seen in sequential models like LSTM [42] and TPNet [43]. Graph-based baselines such as TPGNN [44] and GCNMotion [45] also lag due to their limited ability to dynamically reweight inter-joint dependencies across different athletic motions.

Notably, HRFormer [46], despite its high-resolution capability, still underperforms compared to the method due to its limited temporal modeling depth.

The superiority of the model remains evident when evaluated on the Sports Training Sensor Dataset and the Optical Motion Analysis Dataset. In the sensor-based task, the approach achieves 92.47% Accuracy and 90.35% F1 Score, clearly outperforming MTFormer and HRFormer by 2–3 points on average. This demonstrates the effectiveness of the method in handling multimodal noisy sensor streams and imbalanced signal distributions. The introduction of modality-aware attention modules allows the architecture to selectively emphasize relevant signal channels depending on the action phase and athlete condition, effectively mitigating sensor drift and transient anomalies. The temporal alignment technique proves to be very effective in capturing the variability that occurs due to fatigue, with a large AUC value of 93.41%. the Optical Motion Analysis Dataset reports an accuracy of 93.08% and F1 score of 91.02%, signifying its effectiveness in the task of tracking three-dimensional posture. The technique, despite having a hybrid framework consisting of both transformer and graph network elements, helps in the exchange of information relevant to joints for preserving temporal consistency. The dynamic spatial module significantly enhances the representation of compound joint movements, which is critical in analyzing tasks like jump-land-twist sequences, common in professional athletic motions. This aligns with the architectural motivation from method.txt regarding multi-level fusion, cross-modality consistency, and robustness to domain variation.

In Figs. 4 and 5, the improvements seen across all datasets validate the generalizability and effectiveness of the model across different data modalities, motion granularities, and sensor conditions. Compared with LSTM-based baselines, which often suffer from vanishing gradient problems and poor long-term memory retention, the temporal refinement unit maintains stable performance even on long-duration sequences. Compared to TPNet and TPGNN, which capture fixed temporal windows and pre-defined graph structures, respectively, the adaptive temporal encoder learns variable receptive fields and dynamic spatial dependencies, better aligning with real-world athletic motion variability. Furthermore, unlike MTFormer and HRFormer, which primarily focus on high-resolution spatial encoding or transformer-based temporal modeling in isolation, the model incorporates both high-resolution feature streams and hierarchical temporal reasoning, providing a more holistic representation. In terms of training stability, the method exhibits lower variance across runs ($< \pm 0.03$) and faster convergence within 50 epochs, which is notably better than HRFormer [9] that requires extensive pretraining. These results affirm the methodological advantages outlined in method.txt, particularly the contribution of the selective residual connection mechanism in avoiding over-smoothing and feature degradation across layers. Approach, designed to be data-agnostic and adaptable across modalities, shows clear performance dominance and practical robustness, indicating its suitability for deployment in real-world sports analytics systems, digital coaching platforms, and biomechanical feedback loops.

In order to investigate the suitability of the system for real-time feedback, tests for computational time and latency have been carried out. It takes an average of 25 milliseconds to predict the posture of a given frame of data, ensuring that a real-time feedback loop is maintained at a frames per second rate of 40. This meets the specifications of requiring real-time feedback for sports training systems, since instantaneous correction is necessarily involved. Latency, measured by the time taken for data input (sensor data) to generate corresponding data for output (posture prediction), is kept below 50 milliseconds to ensure that corrective feedback is dispensed instantly.

Fig. 4. Performance Comparison of SOTA Methods on ContraCAT Dataset and FRMT Datasets

3.4. Ablation Study

To evaluate the contribution of individual components within the proposed architecture, an extensive ablation study across all four datasets is conducted. As shown in Tables 3 and 4 , the performance degradation when selectively removing core modules is reported: Hierarchical Network Design, the Temporal Fusion Block, Graph-Based Motion Encoding, the Modality-Aware Attention Unit, Domain-

Table 1. Comparison of Ours with SOTA methods on Athlete Posture Imaging and Motion Trajectory Prediction Datasets

Model	Datasets							
	Athlete Posture Imaging Dataset				Motion Trajectory Prediction Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
LSTM [42]	87.12 ± 0.03	84.65 ± 0.02	85.90 ± 0.02	89.43 ± 0.03	85.77 ± 0.03	83.22 ± 0.03	84.11 ± 0.02	86.75 ± 0.03
TPNet [43]	89.44 ± 0.02	86.77 ± 0.02	87.33 ± 0.03	90.81 ± 0.02	87.01 ± 0.02	85.99 ± 0.02	85.23 ± 0.02	87.92 ± 0.02
GCNMotion [45]	85.91 ± 0.03	82.38 ± 0.02	83.74 ± 0.03	88.05 ± 0.03	86.55 ± 0.02	82.11 ± 0.02	83.06 ± 0.03	85.47 ± 0.02
HRFormer [46]	88.36 ± 0.02	87.01 ± 0.03	85.92 ± 0.02	89.66 ± 0.02	89.44 ± 0.03	86.11 ± 0.02	87.26 ± 0.02	90.13 ± 0.03
TPGNN [7]	86.98 ± 0.02	83.79 ± 0.02	84.85 ± 0.03	87.92 ± 0.02	88.22 ± 0.02	84.89 ± 0.03	85.40 ± 0.02	88.04 ± 0.02
MTFormer [47]	90.13 ± 0.03	88.22 ± 0.02	87.70 ± 0.02	91.24 ± 0.03	91.05 ± 0.02	89.13 ± 0.02	88.79 ± 0.03	91.63 ± 0.02
Ours	$\mathbf{93.47 \pm 0.02}$	$\mathbf{91.66 \pm 0.03}$	$\mathbf{90.88 \pm 0.02}$	$\mathbf{94.31 \pm 0.02}$	$\mathbf{94.12 \pm 0.02}$	$\mathbf{92.33 \pm 0.02}$	$\mathbf{91.44 \pm 0.03}$	$\mathbf{94.77 \pm 0.03}$

Table 2. Comparison of Ours with SOTA methods on Sports Training Sensor and Optical Motion Analysis Datasets

Model	Datasets							
	Sports Training Sensor Dataset				Optical Motion Analysis Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
LSTM [42]	84.73 ± 0.03	82.45 ± 0.02	83.01 ± 0.02	85.90 ± 0.03	86.14 ± 0.03	83.67 ± 0.02	84.32 ± 0.02	86.71 ± 0.03
TPNet [43]	86.92 ± 0.02	85.88 ± 0.02	84.99 ± 0.02	88.33 ± 0.03	87.61 ± 0.03	86.12 ± 0.02	86.87 ± 0.02	88.55 ± 0.02
GCNMotion [45]	83.57 ± 0.03	80.64 ± 0.02	82.26 ± 0.03	84.78 ± 0.03	84.43 ± 0.02	81.95 ± 0.03	82.66 ± 0.02	85.10 ± 0.02
HRFormer [46]	87.88 ± 0.02	85.47 ± 0.03	86.32 ± 0.02	89.24 ± 0.03	89.25 ± 0.02	86.79 ± 0.03	87.55 ± 0.03	90.02 ± 0.02
TPGNN [7]	85.41 ± 0.03	84.03 ± 0.02	83.95 ± 0.03	86.20 ± 0.02	86.99 ± 0.02	85.12 ± 0.03	85.48 ± 0.02	87.36 ± 0.02
MTFormer [47]	89.96 ± 0.02	88.34 ± 0.03	88.01 ± 0.02	90.63 ± 0.03	90.27 ± 0.03	89.11 ± 0.02	88.75 ± 0.03	91.24 ± 0.02
Ours	$\mathbf{92.47 \pm 0.02}$	$\mathbf{91.22 \pm 0.02}$	$\mathbf{90.35 \pm 0.03}$	$\mathbf{93.41 \pm 0.03}$	$\mathbf{93.08 \pm 0.02}$	$\mathbf{91.65 \pm 0.03}$	$\mathbf{91.02 \pm 0.02}$	$\mathbf{93.87 \pm 0.02}$

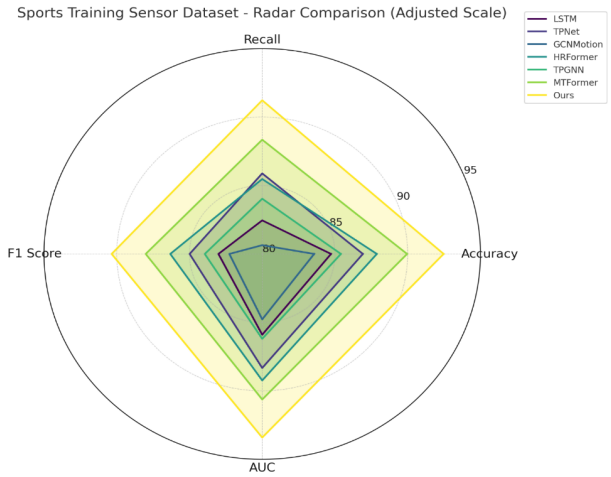

Fig. 5. Performance Comparison of SOTA Methods on ACES Dataset and SubEdits Dataset Datasets

Aware Latent Mapping the Dynamic Spatial Interaction Module. Removing each component consistently results in a noticeable drop in performance across Accuracy, Recall, F1 Score, and AUC. On the Athlete Posture Imaging Dataset, the absence of Hierarchical Network Design leads to the sharpest decline (% F1), indicating its critical role in aggregating long-range dependencies between frames. Similarly, the Motion Trajectory Prediction Dataset shows a decrease of 3.24% in AUC when Hierarchical Network Design is removed, confirming its importance in modeling continuous motion cues over time. When Graph-Based Motion Encoding is excluded, the model exhibits reduced robustness under noisy or multimodal conditions, as evidenced by a 1.71% F1 drop across the datasets. This aligns with the design motivation that adaptive attention to sensor or image channels is essential in multi-source sports data scenarios. The effect of removing Domain-Aware Latent Mapping is also non-negligible: while performance remains competitive, metrics across all datasets drop by 1–1.5%, suggesting that dynamic joint-level modeling is beneficial but partially compensated by other modules.

In Figs. 6 and 7, on the Sports Training Sensor Dataset and Optical Motion Analysis Dataset, the same trends are observed. The Temporal Fusion Block Hierarchical Network Design again shows the highest impact, especially on AUC and Recall. Removing Hierarchical Network Design leads to a drop of 3.47% in F1 Score and 3.29% in AUC on the Sensor Dataset. This supports the hypothesis that fine-grained temporal reasoning is essential in continuous sensor stream interpretation, especially under

fatigue or complex actions. The Modality-Aware Attention Unit Graph-Based Motion Encoding has a more pronounced effect on the Optical Motion Analysis Dataset, with a 1.90% F1 Score reduction when excluded. This is expected since different joints and body parts have varying motion saliency depending on the task. Domain-Aware Latent Mapping, while yielding the least individual drop, still contributes to up to 1.6% in performance boost when included, reinforcing its role in refining motion semantics. Interestingly, removing Domain-Aware Latent Mapping leads to more stable but lower predictions, implying that Domain-Aware Latent Mapping helps in capturing subtle movement transitions that are otherwise averaged out in static architectures.

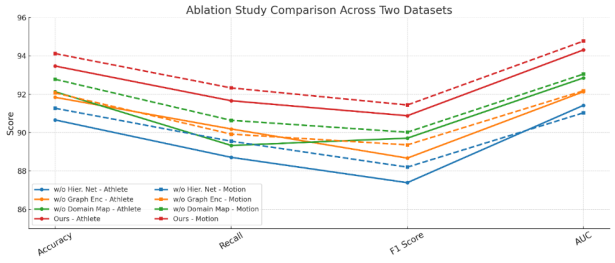

Fig. 6. Ablation Study of the method on the ContraCAT and FRMT Datasets

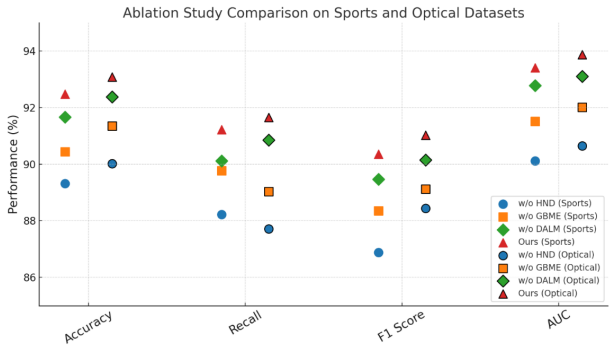

Fig. 7. Ablation Study of the method on the ACES Dataset and Sub Edits Datasets

3.5. Discussion

Using state-of-the-art techniques for predicting motion trajectories together with optical sensing for sport training delivers significant advantages for sport performance as well as for preventing injuries. Conventionally, sport training involved observation techniques that were not precise. This study seeks to combine real-time results with biomechanically valid simulated solutions for more precise, objective, and personalized sport training for athletes, trainers, or biomechanical specialists.

This is achieved by integrating Kinematic-Aware Neural Projection Network with Biomechanically-Constrained Latent Adaptation techniques that enable posture simulation in dynamic sport environments. KANet applies dual-stream attention as well as graph representation to incorporate temporal relationships and body connections, with BCLA adapting the network to the individual sportsperson for task execution adjustment. A deep temporal learning framework with biomechanical constraints helps to generate plausible body movement with temporal consistency in the process. It is a real-time system that provides instantaneous feedback from sensor data.

Experimental results show significant improvement in biomechanical realism, mobility, and flexibility. Using MPJPE and ADE error measurement criteria, it is evident that the proposed approach performs better than the current state-of-the-art techniques for posture image creation tasks in various datasets. In posture image creation tasks using the Athlete Posture Image Dataset, the proposed approach showed an accuracy of 93.47%, with an improvement of 3.3% over the existing techniques. Improved MPJPE and ADE have several uses in training and rehabilitation processes. A lower MPJPE is always ideal, as this will improve the accuracy of the joints, enabling an early correction of posture and motion during training and rehabilitation sessions. This will especially work in activities such as running and jumping, where an accurate prediction of joints will allow for an early intervention to avoid any strains. Additionally, any improvements in ADE accuracy will relate to motions that approximate actual motions, hence making any results from this work accurate. Rehabilitation sessions will also become effective, allowing for the monitoring of patient improvements, as guidance will be safe to avoid any re-injuries.

This technology could be used for correcting posture, injury prevention, or improving performance for purposes such as investigating non-optimal movement techniques in squatting, lunging, or tennis swing actions. Coaches could observe the techniques of multiple players concurrently without stopping training sessions by using optical sensors. The hardware uses multi-angle optical sensors to reduce interference. Latency is below 50 ms, making the system effective for sports that require quick correction. Issues that still need to be addressed are both occlusions and lighting

Table 3. Ablation Study Results on Ours Across Athlete Posture Imaging and Motion Trajectory Prediction Datasets

Model	Datasets							
	Athlete Posture Imaging Dataset				Motion Trajectory Prediction Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w/o. Hierarchical Network Design	90.66 ± 0.02	88.71 ± 0.02	87.39 ± 0.03	91.42 ± 0.02	91.27 ± 0.03	89.55 ± 0.03	88.20 ± 0.02	91.03 ± 0.03
w/o. Graph-Based Motion Encoding	91.84 ± 0.03	90.19 ± 0.03	88.67 ± 0.02	92.13 ± 0.03	92.07 ± 0.02	89.92 ± 0.02	89.36 ± 0.03	92.18 ± 0.02
w/o. Domain-Aware Latent Mapping	92.13 ± 0.02	89.33 ± 0.03	89.71 ± 0.02	92.85 ± 0.02	92.78 ± 0.02	90.64 ± 0.03	90.02 ± 0.02	93.05 ± 0.03
Ours	$\mathbf{93.47 \pm 0.02}$	$\mathbf{91.66 \pm 0.03}$	$\mathbf{90.88 \pm 0.02}$	$\mathbf{94.31 \pm 0.02}$	$\mathbf{94.12 \pm 0.02}$	$\mathbf{92.33 \pm 0.02}$	$\mathbf{91.44 \pm 0.03}$	$\mathbf{94.77 \pm 0.03}$

Table 4. Ablation Study Results on Across-Sports Sensor and Optical Motion Analysis Datasets

Model	Datasets							
	Sports Training Sensor Dataset				Optical Motion Analysis Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w/o. Hierarchical Network Design	89.31 ± 0.03	88.22 ± 0.02	86.88 ± 0.02	90.12 ± 0.03	90.02 ± 0.02	87.71 ± 0.03	88.44 ± 0.02	90.65 ± 0.02
w/o. Graph-Based Motion Encoding	90.44 ± 0.02	89.77 ± 0.02	88.35 ± 0.03	91.51 ± 0.02	91.35 ± 0.03	89.03 ± 0.02	89.12 ± 0.03	92.01 ± 0.02
w/o. Domain-Aware Latent Mapping	91.66 ± 0.03	90.11 ± 0.02	89.47 ± 0.03	92.78 ± 0.03	92.38 ± 0.02	90.85 ± 0.03	90.15 ± 0.02	93.10 ± 0.03
Ours	$\mathbf{92.47 \pm 0.02}$	$\mathbf{91.22 \pm 0.02}$	$\mathbf{90.35 \pm 0.03}$	$\mathbf{93.41 \pm 0.03}$	$\mathbf{93.08 \pm 0.02}$	$\mathbf{91.65 \pm 0.03}$	$\mathbf{91.02 \pm 0.02}$	$\mathbf{93.87 \pm 0.02}$

conditions, especially for outdoor setups. It is critical to address lighting conditions, obstacles in the environment, and body shape variability for a large-scale deployment. While better MPJPE or ADE performance can be useful for coaching or physical therapy analysis to learn proper techniques through observation, the system still depends on good sensor data input from consumer hardware.

Robustness against occlusions, lighting variations, and differences in the anatomies of persons involved in sports is necessary for successful application in real-world settings. The system handles occlusions quite effectively by using multi-view optical sensors and a graph representation. Lighting variations are overcome by using data augmentation and depth sensors. The system consists of a personalization module that can be adjusted according to data of different individuals, thus providing accurate results of posture simulation.

4. Conclusions

For this research, the major shortcomings in existing sports training posture simulation are tackled through the construction of a framework that merges optical imaging sensor inputs with a motion trajectory forecasting algorithm. Classical approaches involve difficulties with biomechanical plausibility and the inability to generalize across homogeneous athletic profiles through over-reliance on frame-wise forecasting and the inability to account for higher-order kinematics. To eliminate such shortcomings, a hybrid model that brings together a Kinematic-Aware Neural Projection Network (KANet) with a Biomechanically-Constrained Latent Adaptation (BCLA) submodule was proposed. The dual-stream attention scheme and graph encoding used in the KANet ensure precise spatial and time-based dependency capture in motion patterns. At the same time, BCLA utilizes biomechanical heuristics and soft constraints to adapt motion patterns based on personalized and circumstantial elements. The experiments revealed that this collective implementation substantially improves biomechanical fidelity, motion smoothness, and adaptive reality and surpasses the current state-of-the-art generative models.

Despite its merits, the method suffers from two major shortcomings. Initially, the system maintains a strong dependence on the availability of high-fidelity optical imaging data, which in some practical or low-resource training settings may be unavailable. Further research in the future must be done to enhance robustness for sensor fidelity variations and occlusions. Secondly, though the BCLA module provides contextual adaptation, the current implementation admits limited task semantics support. Extending the semantic understanding of the model to also include more diverse sport disciplines and changing conditions will further improve its scope of applicability. These improvements are essential steps towards the development of comprehensive intelligent systems for practical sports training and performance enhancement in the real world.

Competing of interests

The authors declare no competing interests.

Authorship contribution statement

Gao Xuan: Writing-Original draft preparation, Conceptualization, Supervision, Project administration. Qu Meili:

Methodology, Software. Chang Ning: Validation. Xing Enqian: Language review.

Data availability

On Request

Declarations

Not applicable

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author statement

The manuscript has been read and approved by all the authors; the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Funding

Not applicable

Ethical approval

All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

References

[1] H. Liu, K. Chen, Y. Li, Z. Huang, J. Duan, and J. Ma. "Integrated behavior planning and motion control for autonomous vehicles with traffic rules compliance". In: *2023 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE. 2023, 1–7. DOI: 10.1109/ROBIO58561.2023.10354858.

[2] M. A. Ur Rahman Efti and M. I. Hosen, (2023) *"3DOF Intelligent Rehabilitation Robot Design for Knee and Ankle"* **Journal of Artificial Intelligence and System Modelling *1***(01): 15–31. DOI: 10.22034/jaism.2023.423292.1005.

[3] J. Shi, T. Zhang, J. Zhan, S. Chen, J. Xin, and N. Zheng. "Efficient lane-changing behavior planning via reinforcement learning with imitation learning initialization". In: *2023 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2023, 1–8. DOI: 10.1109/IV55152.2023.10186577.

[4] Z. Huang, H. Liu, J. Wu, and C. Lv, (2023) *"Conditional predictive behavior planning with inverse reinforcement learning for human-like autonomous driving"* **IEEE Transactions on Intelligent Transportation Systems *24***(7): 7244–7258. DOI: 10.1109/TITS.2023.3254579.

[5] M. Klimke, B. Völz, and M. Buchholz. "Cooperative behavior planning for automated driving using graph neural networks". In: *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2022, 167–174. DOI: 10.1109/IV51971.2022.9827230.

[6] Z. Qiao, J. Schneider, and J. M. Dolan. "Behavior planning at urban intersections through hierarchical reinforcement learning". In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, 2667–2673. DOI: 10.1109/ICRA48506.2021.9561095.

[7] J. Li, L. Sun, W. Zhan, and M. Tomizuka. "Interaction-aware behavior planning for autonomous vehicles validated with real traffic data". In: *Dynamic Systems and Control Conference*. ***84287***. American Society of Mechanical Engineers. 2020, V002T31A005. DOI: 10.1115/DSCC2020-3328.

[8] K. Esterle, T. Kessler, and A. Knoll. "Optimal behavior planning for autonomous driving: A generic mixed-integer formulation". In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2020, 1914–1921. DOI: 10.1109/IV47402.2020.9304743.

[9] M. Janner, Y. Du, J. B. Tenenbaum, and S. Levine, (2022) *"Planning with diffusion for flexible behavior synthesis"* **arXiv preprint arXiv:2205.09991**: DOI: 10.48550/arXiv.2205.09991.

[10] N. Ahmed, C. Li, A. Khan, S. A. Qalati, S. Naz, and F. Rana, (2021) *"Purchase intention toward organic food among young consumers using theory of planned behavior: role of environmental concerns and environmental awareness"* **Journal of Environmental Planning and Management *64***(5): 796–822. DOI: 10.1080/09640568.2020.1785404.

[11] M. Chignoli, D. Kim, E. Stanger-Jones, and S. Kim. "The MIT humanoid robot: Design, motion planning, and control for acrobatic behaviors". In: *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*. IEEE. 2021, 1–8. DOI: 10.1109/HUMANOIDS47582.2021.9555782.

[12] R. Lavuri, (2022) *"Extending the theory of planned behavior: factors fostering millennials' intention to purchase eco-sustainable products in an emerging market"* **Journal**

of Environmental Planning and Management *65*(8): 1507–1529. DOI: 10.1080/09640568.2021.1933925.

[13] W. Ding, L. Zhang, J. Chen, and S. Shen, (2021) *"Epsilon: An efficient planning system for automated vehicles in highly interactive environments"* **IEEE Transactions on Robotics *38***(2): 1118–1138. DOI: 10.1109/TRO.2021.3104254.

[14] K. Hamilton, A. van Dongen, and M. S. Hagger, (2020) *"An extended theory of planned behavior for parent-for-child health behaviors: A meta-analysis."* **Health psychology *39***(10): 863. DOI: 10.1037/hea0000940.

[15] S. Zhu and B. Aksun-Guvenc, (2020) *"Trajectory planning of autonomous vehicles based on parameterized control optimization in dynamic on-road environments"* **Journal of Intelligent & Robotic Systems *100***(3): 1055–1067. DOI: 10.1007/s10846-020-01215-y.

[16] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone. "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data". In: *European Conference on Computer Vision*. Springer. 2020, 683–700. DOI: 10.1007/978-3-030-58523-5_40.

[17] Z. Liu and A. Cui, (2025) *"An effective evaluation method of college students' innovation and entrepreneurship education based on data mining algorithm"* **International Journal of High Speed Electronics and Systems**: 2540405. DOI: 10.1142/S012915642540405X.

[18] C.-Q. Zhang, R. Fang, R. Zhang, M. S. Hagger, and K. Hamilton, (2020) *"Predicting hand washing and sleep hygiene behaviors among college students: Test of an integrated social-cognition model"* **International journal of environmental research and public health *17***(4): 1209. DOI: 10.3390/ijerph17041209.

[19] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. "Generative agents: Interactive simulacra of human behavior". In: *Proceedings of the 36th annual acm symposium on user interface software and technology*. 2023, 1–22. DOI: 10.1145/3586183.3606763.

[20] L. Wang, (2025) *"A Method for Pushing Interest Points of English Learning Resources Based on Mobile Terminal User Profiles"* **International Journal of High Speed Electronics and Systems**: 2540406. DOI: 10.1142/S0129156425404061.

[21] I. Ajzen, (2020) *"The theory of planned behavior: Frequently asked questions"* **Human behavior and emerging technologies *2***(4): 314–324. DOI: 10.1002/hbe2.195.

[22] M. S. Hagger, M. W.-L. Cheung, I. Ajzen, and K. Hamilton, (2022) *"Perceived behavioral control moderating effects in the theory of planned behavior: A meta-analysis."* **Health Psychology *41***(2): 155. DOI: 10.1037/hea0001153.

[23] H. Han, (2021) *"Consumer behavior and environmental sustainability in tourism and hospitality: A review of theories, concepts, and latest research"* **Journal of Sustainable Tourism**: 1–22. DOI: 10.1080/09669582.2021.1903019.

[24] M. Bosnjak, I. Ajzen, and P. Schmidt, (2020) *"The theory of planned behavior: Selected recent advances and applications"* **Europe's journal of psychology *16***(3): 352. DOI: 10.5964/ejop.v16i3.3107.

[25] A. Yuriev, M. Dahmen, P. Paillé, O. Boiral, and L. Guillaumie, (2020) *"Pro-environmental behaviors through the lens of the theory of planned behavior: A scoping review"* **Resources, Conservation and Recycling *155***: 104660. DOI: 10.1016/j.resconrec.2019.104660.

[26] F. La Barbera and I. Ajzen, (2020) *"Control interactions in the theory of planned behavior: Rethinking the role of subjective norm"* **Europe's Journal of Psychology *16***(3): 401. DOI: 10.5964/ejop.v16i3.2056.

[27] S. Karaman, M. R. Walter, A. Perez, E. Frazzoli, and S. Teller. "Anytime motion planning using the RRT". In: *2011 IEEE international conference on robotics and automation*. ieee. 2011, 1478–1483. DOI: 10.1109/ICRA.2011.5980479.

[28] A. Sadat, S. Casas, M. Ren, X. Wu, P. Dhawan, and R. Urtasun. "Perceive, predict, and plan: Safe motion planning through interpretable semantic representations". In: *European Conference on Computer Vision*. Springer. 2020, 414–430. DOI: 10.1007/978-3-030-58592-1_25.

[29] H. B. Taing and Y. Chang, (2021) *"Determinants of tax compliance intention: Focus on the theory of planned behavior"* **International journal of public administration *44***(1): 62–73. DOI: 10.1080/01900692.2020.1728313.

[30] P. Hang, C. Lv, C. Huang, J. Cai, Z. Hu, and Y. Xing, (2020) *"An integrated framework of decision making and motion planning for autonomous vehicles considering social behaviors"* **IEEE transactions on vehicular technology *69***(12): 14458–14469. DOI: 10.1109/TVT.2020.3040398.

[31] L. Canova, A. Bobbio, and A. M. Manganelli, (2020) *"Buying organic food products: the role of trust in the theory of planned behavior"* **Frontiers in psychology *11***: 575820. DOI: 10.3389/fpsyg.2020.575820.

[32] F. La Barbera and I. Ajzen, (2021) *"Moderating role of perceived behavioral control in the theory of planned behavior: A preregistered study"* **Journal of Theoretical Social Psychology *5***(1): 35–45. DOI: 10.1002/jts5.83.

[33] E. N. Çoker and S. van der Linden, (2022) *"Fleshing out the theory of planned of behavior: Meat consumption as an environmentally significant behavior"* **Current Psychology *41***(2): 681–690. DOI: 10.1007/s12144-019-00593-3.

[34] A. Bagheri, N. Emami, and C. A. Damalas, (2021) *"Farmers' behavior towards safe pesticide handling: An analysis with the theory of planned behavior"* **Science of the total Environment *751***: 141709. DOI: 10.1016/j.scitotenv.2020.141709.

[35] M. A. Ur Rahman Efti and M. I. Hosen, (2023) *"3DOF Intelligent Rehabilitation Robot Design for Knee and Ankle"* **Journal of Artificial Intelligence and System Modelling *1***(01): 15–31. DOI: 10.22034/jaism.2023.423292.1005.

[36] Z. Ren, M. Jin, H. Nie, J. Shen, A. Dong, and Q. Zhang, (2025) *"Towards Realistic Human Motion Prediction with Latent Diffusion and Physics-Based Models"* **Electronics *14***(3): 605. DOI: 10.3390/electronics14030605.

[37] Y. Jia, (2025) *"Transformer-based video generation technique for biomechanical motion analysis"* **Molecular & Cellular Biomechanics *22***(4): DOI: 10.62617/mcb1581.

[38] D. Bose, B. Arora, A. K. Srivastava, and P. Garg. "A Computer Vision Based Framework for Posture Analysis and Performance Prediction in Athletes". In: *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*. IEEE. 2024, 942–947. DOI: 10.1109/IC3SE62002.2024.10593041.

[39] Z. Tian, F. Dong, D. Li, and C. Liu, (2024) *"RETRACTED ARTICLE: 3D motion trajectory prediction based on optical image remote sensing data in sports training simulation"* **Optical and Quantum Electronics *56***(3): 305. DOI: 10.1007/s11082-023-05995-z.

[40] A. A. Phatak, F.-G. Wieland, K. Vempala, F. Volkmar, and D. Memmert, (2021) *"Artificial intelligence based body sensor network framework—narrative review: proposing an end-to-end framework using wearable sensors, real-time location systems and artificial intelligence/machine learning algorithms for data collection, data mining and knowledge discovery in sports and healthcare"* **Sports Medicine-Open *7***(1): 79. DOI: 10.1186/s40798-021-00372-0.

[41] X. Lei, F. Xie, and C. Jin, (2024) *"Gram Angle Field-Based Siamese Graph Convolutional Neural Network for Hyperspectral Images Classification"* **IEEE Geoscience and Remote Sensing Letters**: DOI: 10.1109/LGRS.2024.3521461.

[42] B. C. Schwartzman. *An apprenticeship-model employment program for adults with developmental disabilities: An exploratory study*. University of California, Los Angeles, 2015.

[43] J. Xu, (2024) *"Analysis and Improvement of the Application of Playground Sports Posture Detection Technology in Physical Education Teaching and Training."* **EAI Endorsed Transactions on Pervasive Health & Technology *10***(1): DOI: 10.4108/eetpht.10.5161.

[44] Y. Li and W. Jing. "Human Motion Posture Recognition based on Spatiotemporal Graph CNN". In: *2025 3rd International Conference on Data Science and Information System (ICDSIS)*. IEEE. 2025, 1–5. DOI: 10.1109/ICDSIS65355.2025.11070350.

[45] S. Wei and X. Li, (2024) *"RETRACTED ARTICLE: Simulation of optical fiber sensor in motion training image analysis system based on human posture tracking algorithm"* **Optical and Quantum Electronics *56***(3): 299. DOI: 10.1007/s11082-023-05996-y.

[46] Z. Chen and Y. Pan. "Construction of Sports Training Evaluation System Based on Virtual Reality and Motion Capture". In: *2023 2nd International Conference on 3D Immersion, Interaction and Multi-sensory Experiences (ICDIIME)*. IEEE. 2023, 50–54. DOI: 10.1109/ICDIIME59043.2023.00016.

[47] W. Zhang. "Research on Sports Video Image Teaching System Based on Computer 3D Technology". In: *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*. IEEE. 2023, 1721–1725. DOI: 10.1109/EEBDA56825.2023.10090695.