Prediction and Optimization of Motion Trajectory of Hybrid Mechanism Based on Transformer

Haofeng Cao[1]
[1]Teaching Center, Ningbo Open University, Ningbo 315016, Zhejiang, China
Corresponding author: Haofeng Cao (e-mail: haofeng_cao@163.com)

ABSTRACT To address error accumulation from multi-variable features and complex dynamic coupling in the motion trajectory prediction of hybrid mechanisms, this paper applies a Transformer-based prediction framework. The framework integrates a spatiotemporal attention mechanism and kinematic constraints to obtain precise and physically feasible predictions of component movement. Such trajectory modeling is also meaningful for advanced electromagnetic engineering, including antenna-positioning mechanisms, microwave inspection platforms, and robotic systems used in electromagnetic measurement. The method uses a Transformer core with a hierarchical spatiotemporal attention module to model time dependency and spatial coupling among joint angles, velocities, and torques, while refining multimodal dynamic behavior. A Dynamic Graph Convolutional Network captures the influence of topological changes on motion-transfer paths and reflects configuration-dependent joint interactions. By constructing a kinematic constraint loss function based on the Lagrangian equation, the model embeds dynamic priors and improves physical consistency. Experimental results show that the method achieves an RMSE of 1.42 mm and an MAE of 1.06 mm on the standard test set, reducing these errors by 63.3% and 64.1% compared with a traditional LSTM model. In a six-degreeof-freedom mechanism, the median prediction error of critical Joint 1 is controlled at 1.42 mm, supporting accurate motion prediction for constrained serial-reachable hybrid topologies.

INDEX TERMS trajectory prediction,transformer,dynamic graph convolution,kinematic constraints,antenna positioning mechanism

I. INTRODUCTION

As a complex mechanical system combining the advantages of parallel and series structures, the hybrid mechanism has a wide range of applications in high-precision operations and multi-degree-of-freedom motion control tasks. Due to its flexible topological structure and high degree of kinematic coupling, the hybrid mechanism has higher requirements in trajectory generation and control accuracy [1-3]. The end effector of the hybrid mechanism is affected by the motion states of multiple joints. Its trajectory behavior is not only affected by the current state variables but also depends on the historical state sequence and the dynamic response of the structural transmission path [4-6].

In practical applications, hybrid mechanisms often involve multi-modal input features, such as joint angles, angular velocities, joint torques, etc. There are complex spatiotemporal interactions between these state variables, making it difficult to obtain accurate trajectory prediction results directly through static mapping modeling [7,8]. Traditional analytical modeling methods need to rely on a large amount of manual derivation during modeling. When faced with high-dimensional, multi-constrained, and nonlinear strongly coupled systems, it is difficult to maintain modeling accuracy and generalization capabilities [9,10]. Although trajectory prediction methods based on RNN (Recurrent Neural Network) and their variants have certain advantages in processing time series data, they are unstable in complex time series trajectories due to their weak ability to model long-term dependencies and serious gradient dissipation [11-13]. This type of model generally treats each state variable as an independent sequence, ignoring the spatial transmission path within the hybrid structure and the structural relationship between nodes, making it difficult to reflect the impact of the mechanism topology on the motion behavior [14,15].

When fusing multi-modal input features, existing methods often adopt shared weights or direct concatenation strategies in the temporal modeling process without considering the changes in the correlation strength between different modal variables, resulting in information redundancy or coupling imbalance problems in prediction [16-18]. Since data-driven models generally lack physical prior constraints, it is difficult to effectively avoid the emergence of non-physical solutions, which leads to a decrease in the prediction accuracy of the model when facing complex load disturbances or boundary state changes, and causes error accumulation problems [19,20]. How to construct a prediction framework that takes into account the decoupling of multi-modal dynamic behavior, structural topology modeling capabilities, and physical consistency has become a key challenge in the motion trajectory modeling of hybrid mechanisms [21,22]. To address the challenges of error accumulation and dynamic coupling in complex machinery, this paper proposes a novel Transformer trajectory prediction framework that integrates a spatiotemporal at-tention mechanism and kinematic constraints. This constitutes a comprehensive motion modeling strategy driven by multi-modal input decoupling, structural perception, and strict physical constraints.

II. RELATED WORK

In the study of hybrid mechanism motion modeling and trajectory prediction, the integrated application of optimization algorithms and control strategies has become a key direction for improving trajectory accuracy and system performance. Singh et al. used a method combining seventh-order polynomials with hybrid metaheuristic optimization algorithms to optimize the joint trajectory design of three industrial robots. By constraining jerk, acceleration, and velocity parameters to improve motion smoothness, the experiments showed that the proposed scheme was superior to existing methods in shortening travel time and obstacle avoidance performance [23]. On the basis of completing the optimization of motion smoothness, the research further focused on improving the efficiency of the algorithm and constructing an optimization strategy for global path performance [24,25]. Singh et al. combined 18 optimization algorithms into a hybrid algorithm to perform forward and inverse kinematics analysis and trajectory planning on three industrial robots. Simulation verification showed that the hybrid algorithm significantly shortened the calculation time and number of iterations, and achieved the shortest travel time for each robot [26]. While improving trajectory planning performance, the research focus has gradually extended to the control robustness and system dynamic stability issues in the trajectory tracking process [27,28]. Kordi et al. proposed a hybrid control strategy that combines predictive control with finite-time integral terminal sliding mode control. They designed a nonlinear disturbance observer to effectively deal with external disturbances and parameter uncertainties in the trajectory tracking of a wheeled mobile robot, and proved the finite-time stability of the system based on the Lyapunov theory. The simulation results verified the significant advantages of this method in improving tracking accuracy and anti-interference performance [29]. Although these studies have made important progress in path optimization and control robustness, they still lack a unified modeling framework and an adaptive prediction mechanism for complex working conditions, which limits the generalization ability and engineering scalability of the methods.

The Transformer model shows significant potential in time series prediction and optimization control. For instance, Xu et al. used a dual-input Transformer for trajectory prediction, fusing scene and multi-modal interaction data via multi-head attention, which substantially reduced displacement error on the Argoverse dataset [30]. However, even this high-precision method requires further optimization of control mechanisms in complex interactive environments for improved overall feasibility [31,32]. Jiao et al. introduced a hierarchical framework combining deep learning and reinforcement learning, using a Transformer graph neural network to capture multi-scale interactions and generate multi-modal predictions, which was then optimized with vehicle dynamics and reinforcement learning to enhance the feasibility of autonomous driving trajectories in complex scenarios [33]. The advantages of the Transformer are also expanding beyond transportation into industrial process modeling [34,35]. Zhou et al. proposed an improved Transformer model integrating the self-attention mechanism with VAE-calculated KLD and RMSE for nonlinear dynamic process modeling in the petrochemical industry, demonstrating its versatility in industrial soft sensing and time series prediction [36]. While these studies achieved remarkable results, a common limitation remains: insufficient exploration of model generalization, real-time computing efficiency, and cross-dataset robustness, which currently restricts their full deployment in dynamic industrial systems.

III. PREDICTION MODEL CONSTRUCTION AND OPTIMIZATION MECHANISM

The interaction between multi-modal state and structural information determines the prediction performance and physical consistency. To clearly present each functional module and its coupling path, this paper draws the overall architecture diagram of the model shown in Figure 1.

Figure 1 outlines the complete trajectory prediction process, from data collection to error back-propagation. The input features are first time-aligned and normalized before entering the spatiotemporal attention module to extract time-dependency and joint topology information. This is concatenated with structural features from the Dynamic Graph Convolutional Network (DGCN) to form a unified high-dimensional representation. The decoder then outputs the predicted three-dimensional position and posture. The objective function is jointly optimized by a physical consistency loss (derived from Lagrangian dynamics) and a trajectory accuracy loss, ensuring the results are both geometrically accurate and physically rational.

A. MULTI-MODAL STATE INPUT MODELING AND PREPROCESSING

The hybrid mechanism involves the temporal changes and spatial coupling of multi-source state information during the motion process, and the state variables present complex

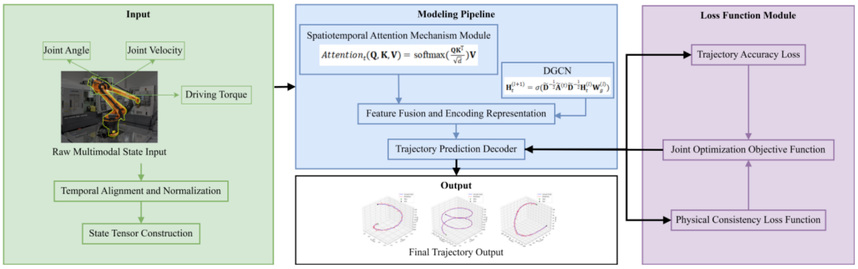

FIGURE 1. Overall architecture of the motion trajectory prediction model of hybrid mechanism

multi-modal dynamic characteristics in the time domain and structural domain. To unify the representation of each state feature and construct a standardized input tensor, the three types of state variables, joint angle, angular velocity and driving torque, need to be regularly encoded according to the time dimension and spatial topology. The state input modeling is unfolded in a time series format, with each trajectory segment length set to T, the structural dimension set to J, and the state variable dimension set to M, where $M = 3$ represents three modes: angle, velocity and torque. The original input data is aligned with the timestamps by linear interpolation to construct a three-dimensional state tensor $X \in \mathbb{R}^{T\times J\times M}$.

To enhance the expressibility of input data, it is necessary to perform standardization and normalization in the state tensor construction stage. Assuming that the original sequence of a modal state variable is $x_{t,j}^{(m)}$, where $t \in [1, T]$, $j \in [1, J]$ and $m \in [1, M]$, which is normalized by formula (1):

$$\hat{x}_{t,j}^{(m)} = \frac{x_{t,j}^{(m)} - \mu^{(m)}}{\sigma^{(m)}} \tag{1}$$

$\mu^{(m)}$ and $\sigma^{(m)}$ in formula (1) represent the mean and standard deviation of the m-th state variable in the training set. Normalization ensures that the numerical scales of different modal variables in the feature space are consistent, avoiding modal bias in gradient propagation. The hierarchical spatiotemporal attention module processes the three modalities after normalization so that torque values do not dominate the attention distribution. The linear embedding layer applies shared scaling across modalities, and the attention computation assigns weights based solely on the learned relevance in the embedded feature space rather than on raw magnitude, ensuring balanced contributions of angle, velocity, and torque.

In terms of state composition, the three types of modal variables are in the form of high-dimensional tensors in spatial structure. Figure 2 shows the spatial distribution of multi-modal state variables of the hybrid mechanism, where the state of each joint node is embedded in the tensor structure in a modal dimension to form a unified input format. The sampling frequency of each state variable is consistent, and the time window sliding acquisition strategy adopts an overlapping window design to maintain the continuity of the input sequence and the stability of the prediction model.

As shown in Figure 2, the multi-modal state input of the hybrid mechanism is uniformly represented as a three-dimensional tensor structure, organized along the time dimension, joint dimension, and modal dimension. This structure vertically stacks three types of state variables on the modal dimension to form three layers of state sub-tensors with the same time length and number of joints, corresponding to angle, velocity, and torque information, respectively. The figure uses red, blue and green to distinguish the modal types, and each small cube represents a state sampling point under a specific time and joint number. The tensor as a whole presents a regular cubic distribution, which vividly reflects the organization of the original state data in the time, space and modal dimensions. This structure provides a unified input format support for subsequent spatiotemporal attention mechanism and topological modeling, enabling the model to achieve fusion modeling of multi-modal dynamic behaviors under a shared structure.

The statistical characteristics of the multi-modal state variables are shown in Table 1, covering the mean, standard deviation, maximum and minimum values of the three core physical quantities in the training set, reflecting their dynamic distribution range and physical fluctuation characteristics.

TABLE 1. Statistical characteristics and modal distribution of state variables

Modal Variable	Physical Quantity	Unit	Mean (μ)	Standard Deviation (σ)	Maximum	Minimum
Modal 1	Joint Angle	deg	47.3	18.5	88.6	3.2
Modal 2	Joint Velocity	deg/s	12.7	6.4	25.1	0.5
Modal 3	Driving Torque	Nm	3.85	1.09	6.42	1.02

The fluctuation range of angle and speed modes is relatively wide, while the distribution of driving torque is relatively concentrated. This feature requires the model to have the ability to distinguish the strength of the mode in the feature extraction stage, avoid the dominant effect of high-fluctuation modes on low-dynamic features and ensure the balanced modeling of input information in the Transformer encoding structure. After preprocessing, the state tensor in a unified format serves as the direct input for subsequent hierarchical attention mechanisms and graph structure modeling, supporting the in-depth analysis and spatial structure modeling of multi-modal dynamic behaviors.

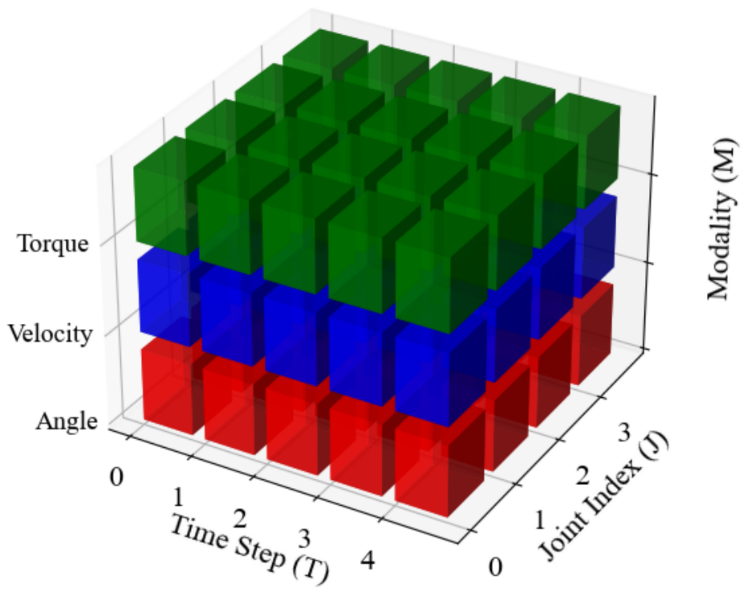

FIGURE 2. Composition of multi-modal state variables of hybrid mechanism

B. HIERARCHICAL SPATIOTEMPORAL ATTENTION MODULE DESIGN

This section constructs a hierarchical spatiotemporal attention module with Transformer as the core to jointly model the time series dependency and spatial coupling relationship in the state input of the hybrid mechanism. The input state tensor is set to $X \in \mathbb{R}^{T \times N \times D}$, where T represents the time step, N represents the number of joints, and D represents the feature dimension of each joint, including joint angle, velocity, and torque. To adapt to the input requirements of the attention structure, The embedding transformation maps all modal inputs into a common latent representation, where their numerical range is harmonized before participating in attention calculation. This prevents torque-related features from being suppressed or amplified due to scale differences and maintains consistent weighting criteria across all modalities. X is first linearly mapped to obtain the embedded representation $H^{(0)} = \mathrm{Linear}(X)$, $H^{(0)} \in \mathbb{R}^{T \times N \times d}$, where d is the embedding dimension.

The temporal attention module takes the time series of a single joint as input and constructs the dependency between each time step. Assuming that the n-th joint is represented by $H_n \in \mathbb{R}^{T \times d}$ in T time steps, its attention is calculated as follows:

$$\mathrm{Attention}_t(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{2}$$

In formula (2), $Q = H_n W_t^Q$, $K = H_n W_t^K$, $V = H_n W_t^V$, and W_t^Q, W_t^K, $W_t^V \in \mathbb{R}^{d \times d}$ are learnable parameters. This process is performed in parallel on all N joints to obtain the temporal attention output tensor $H^{(t)} \in \mathbb{R}^{T \times N \times d}$, which is used to capture the temporal dynamic characteristics of each joint.

The spatial attention module models the dependencies between different joints at the same time step. Taking the t-th time step as an example, the spatial input is represented as $H_t \in \mathbb{R}^{N \times d}$, and its spatial attention calculation form is:

$$\mathrm{Attention}_s(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{3}$$

In formula (3), $Q = H_t W_s^Q$, $K = H_t W_s^K$, $V = H_t W_s^V$, and W_s^Q, W_s^K, $W_s^V \in \mathbb{R}^{d \times d}$ are spatial attention parameters. This process is executed in parallel over T time steps, and the output tensor $H^{(s)} \in \mathbb{R}^{T \times N \times d}$ is used to model the coupling relationship in the spatial structure of the mechanism.

To fuse the two types of attention information, a fusion representation after residual connection and normalization operation is constructed:

$$H^{(l+1)} = \mathrm{LayerNorm}\left(H^{(l)} + \mathrm{FFN}\left(H^{(t)} + H^{(s)}\right)\right) \tag{4}$$

In formula (4), $H^{(l)}$ represents the l-th layer input tensor; FFN represents the feedforward neural network, and the structure is a two-layer fully connected network plus a ReLU activation function to improve the nonlinear modeling ability. The above modules are stacked in L layers to build a deep representation and enhance the ability to express nonlinear dynamics. The hierarchical design produces separate temporal and spatial relevance distributions, and both distributions are computed in the normalized latent domain. This ensures that the torque modality neither overwhelms the encoder nor becomes underrepresented during feature fusion. The final output tensor $H^{(L)} \in \mathbb{R}^{T\times N\times d}$ is fed into the subsequent topology modeling module and trajectory prediction decoder as a joint representation of the multi-modal dynamic state, ensuring that the trajectory output accurately reflects the time dependency and mechanism structure coupling.

C. DGCN INTEGRATION FOR TOPOLOGY MODELING

In this section, the topological structure of the hybrid mechanism is encoded into a dynamic graph neural network structure and combined with DGCN to extract structural dependency information to enhance the model's adaptability to different connection configurations and changes in mechanism morphology. The time-varying adjacency structure expresses changes in the motion transfer pattern caused by configuration evolution, and the physical arrangement of links remains fixed while the graph updates its weights to represent altered transmission influence during operation. Assuming that the topological structure of the hybrid mechanism at time step t is represented by a graph G(t)=(V,E(t)), where V is a set of nodes consisting of N motion units in the mechanism, and E(t) represents the connection relationship between the nodes, which changes dynamically according to the time step. The coupling strength between joints varies with the instantaneous configuration and the transmitted load, which alters the effective transfer direction and intensity along the motion path. This behavior introduces temporal variation in the adjacency representation, so the spatial dependency cannot be described by a fixed structural pattern.

The node feature is represented as a tensor $H^{(L)} \in \mathbb{R}^{T\times N\times d}$, which is output by the spatiotemporal attention module in the previous section. At each time step t, the current graph structure $G(t)$ is extracted to construct an adjacency matrix $A(t) \in \mathbb{R}^{N\times N}$, whose element $a_{ij}^{(t)}$ indicates whether there is a direct motion connection between node i and node j, and the weight $w_{ij}^{(t)}$ calculated based on the transmission path length and degree of freedom constraint is added to the edge, finally forming a weighted adjacency matrix $\tilde{A}^{(t)}$. The adjacency weights follow the configuration-dependent changes in relative joint position and path geometry, which leads to gradual temporal differences in the structural connectivity representation. The input feature of the node in the graph is defined as $H_t \in \mathbb{R}^{N\times d}$, and the graph convolution propagation process is as follows:

$$H_t^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}^{(t)}\tilde{D}^{-\frac{1}{2}}H_t^{(l)}W_g^{(l)}\right) \tag{5}$$

In formula (5), $\tilde{D}$ is the degree matrix; $W_g^{(l)} \in \mathbb{R}^{d\times d'}$ is the weight matrix of the l-th layer graph convolution; and σ is the activation function. This process is performed independently at each time step t to construct the dynamic graph sequence output $Z(t) \in \mathbb{R}^{N\times d'}$, which is finally stacked at all time steps to form a three-dimensional tensor $Z \in \mathbb{R}^{T\times N\times d'}$.

To enhance the sensitivity of graph representation to structural changes, a graph state transfer function is applied to update the adjacency matrix sequence, and the dynamic change rate is calculated based on the structural difference between adjacent time steps:

$$\Delta A^{(t)} = \tanh\left(A^{(t)} - A^{(t-1)}\right) \tag{6}$$

This rate of change is used to adjust the edge weights during graph propagation to form a time-series adaptive adjacency matrix sequence, which enables the model to capture the impact of topological perturbations on the power transmission path and state changes. This adjustment reflects the time-varying evolution of coupling relations across the mechanism, confirming that the motion-transfer characteristics do not remain structurally constant during operation. The final output graph Z is concatenated with the spatiotemporal attention output H through residual connection and feature fusion as the high-order state representation of the subsequent trajectory generation module. This dynamic graph modeling method embeds the actual topological evolution in the hybrid mechanism into the neural network structure, constructs a structure-dependent and perceived motion representation, and effectively improves the model's responsiveness to changes in the transmission structure and its predictive stability.

D. DESIGN OF DYNAMIC CONSTRAINT FUNCTION AND CONSTRUCTION OF OPTIMIZATION TARGET

In order to improve the rationality of trajectory prediction results at the physical level, this section constructs a constraint loss function based on the Lagrangian dynamics principle, and embeds the kinematics and dynamics of the hybrid mechanism into the model optimization process. Assuming that the generalized coordinates of the mechanism are $q(t) \in \mathbb{R}^n$, which represents the angle sequence of n joints, and the predicted output is $\hat{q}(t)$. The generalized coordinates selected for the dynamic constraint are the six active joint angles of the hybrid mechanism, and the corresponding velocities and accelerations are derived through discrete-time differences computed directly from the predicted angle sequence. The velocity tensor is obtained from the first-order difference of consecutive predicted angles divided by the sampling interval, and the acceleration tensor is formed by applying a second-order difference to the velocity sequence so that all derivatives enter the dynamic residual computation in the same prediction space. The constructed Lagrangian-based constraint is implemented as a soft penalty term rather than an enforced hard constraint. Its residual form is computed during each training iteration and added to the overall loss so that the predicted accelerations follow the underlying dynamic relations without explicitly imposing an equality condition at the optimization level. The corresponding angular velocity and angular acceleration are $\dot{\hat{q}}(t)$ and $\ddot{\hat{q}}(t)$, respectively, which are calculated by numerical difference:

$$\dot{\hat{q}}(t) = \frac{\hat{q}(t+\Delta t) - \hat{q}(t)}{\Delta t} \tag{7}$$

$$\ddot{\hat{q}}(t) = \frac{\dot{\hat{q}}(t+\Delta t) - \dot{\hat{q}}(t)}{\Delta t} \tag{8}$$

The Lagrangian dynamic equation of the mechanism can be written as:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = \tau \tag{9}$$

In formula (9), $L = T - V$ is the Lagrangian function; T represents kinetic energy; V represents potential energy; and τ is the generalized force vector. The Lagrange equation residuals constructed in this study do not explicitly include Coriolis force, centrifugal force, and friction terms. This is a simplified dynamic model, intended to provide a basic physical consistency guide for data-driven predictions rather than achieving fully accurate dynamic simulations. This simplification is based on the analysis of the hybrid mechanism under typical textile manufacturing conditions, where inertial and gravity forces dominate. Kinetic energy is defined as:

$$T = \frac{1}{2}\dot{q}^{\top} M(q)\dot{q} \tag{10}$$

Potential energy is defined as:

$$V = \sum_{i=1}^{n} m_i g h_i(q) \tag{11}$$

$M(q) \in \mathbb{R}^{n\times n}$ in formula (10) is the joint mass matrix; m_i in formula (11) is the mass of the i-th component; g is the gravity constant; and $h_i(q)$ represents the center of mass height of the i-th component. Based on the above formula, a physical consistency loss function is constructed to constrain the predicted acceleration and the derived results of the dynamic model. The residual term incorporates the predicted joint angles together with numerically differentiated velocities and accelerations into the mass matrix and potential energy expressions so that the Lagrangian equation is evaluated using quantities generated entirely from the model output at each training iteration. Both sides of the Lagrange equation are rewritten as residual terms:

$$r(t) = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\hat{q}}}\right) - \frac{\partial L}{\partial \hat{q}} - \hat{\tau}(t) \tag{12}$$

The residual quantifies the deviation between the predicted dynamic response and the physically derived expression, and this deviation is penalized in the optimization objective. The training process adjusts the model parameters by reducing this residual, which guides the prediction toward consistency with the dynamic principles while retaining the flexibility of data-driven learning. The physical constraint loss function Lphy is constructed as:

$$L_{\text{phy}} = \frac{1}{T}\sum_{t=1}^{T} \|r(t)\|_2^2 \tag{13}$$

This loss sums the squares of the residuals between the model prediction results and the dynamic equations, minimizing the part that violates the dynamic laws. In addition, to ensure the consistency between the predicted trajectory and the true trajectory, the mean square error loss term is applied:

$$L_{\text{mse}} = \frac{1}{T}\sum_{t=1}^{T} \|\hat{q}(t) - q(t)\|_2^2 \tag{14}$$

The dynamic constraint is not enforced as an equality condition during inference. It is incorporated only as an additional soft term in the training objective, so the model learns a prediction space that adheres to the motion dynamics while maintaining numerical stability and avoiding the infeasibility that arises from strict constraint enforcement. The final joint optimization goal is:

$$L_{\text{total}} = \lambda_1 L_{\text{mse}} + \lambda_2 L_{\text{phy}} \tag{15}$$

In formula (15), λ_1 and λ_2 are weight coefficients that control the balance between physical consistency and trajectory accuracy. Through grid search experiments, it was determined that λ_1 was fixed at 1.0 and λ_2 was fixed at 0.5. This set of parameters achieved the best balance between prediction accuracy and physical plausibility on the validation set. The weight coefficients remained fixed during training and were not dynamically adjusted. This loss function is embedded in the back propagation process during the training phase to guide the model to output prediction results that have both geometric accuracy and dynamic rationality, ensuring the trajectory continuity and stability of the hybrid mechanism under high dynamic and multi-constraint conditions.

IV. EXPERIMENTAL SETUP AND EVALUATION PROTOCOL

A. EXPERIMENTAL PLATFORM AND HARDWARE CONFIGURATION DESCRIPTION

This paper provides a technical description of the experimental implementation platform of the trajectory prediction model, covering the system hardware and software environment, execution unit configuration, and signal transmission scheme. The high degree of freedom and real-time response requirements of the hybrid mechanism determines that the control and simulation environment must have high-precision calculation and high-speed data exchange capabilities. Therefore, the selection and configuration of the experimental platform play a fundamental role in model verification. Combined with the construction requirements and performance goals of the experimental system, Table 2 lists the core platform components and their key parameters.

TABLE 2. Summary of configuration parameters of experimental platform hardware and software environment

Module Category	Configuration Name	Key Parameters	Functional Description
Simulation Platform	MATLAB/Simulink	R2023a, RT (Real-Time) module	Model construction and real-time simulation
Control Unit	NI (National Instruments) CompactRIO-9045	1.91 GHz quad-core, 4 GB RAM	Real-time control execution and data acquisition
Actuator Interface	EtherCAT (Ethernet for Control Automation Technology) Bus Module	Update rate 1 kHz	High-frequency control signal transmission
Sensing System	Optical Encoder + Torque Sensor	Accuracy 0.01° / 0.05 Nm	Acquisition of joint state variables
Operating System	Windows 10 + RT Linux	Dual-system support	Separation of development and deployment environments

The configuration in Table 2 covers the technical parameters of the simulation platform version, real-time control hardware performance indicators, execution layer communication mechanism, and state perception module. The system adopts a dual-platform collaborative operation architecture to ensure a seamless connection between data acquisition and trajectory prediction model. The control end and the sensor end are connected through a high-frequency industrial bus to ensure that the command transmission and response delay are controlled within an acceptable range. The overall architecture emphasizes real-time, stability, and compatibility, which facilitates the deployment and evaluation of the model under various working conditions.

B. DATA COLLECTION AND SAMPLE CONSTRUCTION PROCESS

In the sample construction phase, the state-trajectory supervision pairs are generated by sliding in time series windows. In the preliminary processing stage, the first-order difference smoothing algorithm is used to reduce high-frequency noise interference and maintain the continuity and dynamic trend of state variable changes. All state variables are standardized and the Z-score method is used to complete the normalization transformation based on the mean and standard deviation of the training set to eliminate the impact of scale differences between variables on training stability. The topological structure information is dynamically generated by the adjacency matrix and together with the state tensor, it constitutes the model input. The data set is divided into training set, validation set and test set according to the principle of non-overlapping time, with a ratio of 8:1:1. Each type of data contains all structural forms and state variable distributions to ensure the consistency of evaluation indicators and model generalization ability. The dataset contains samples of various hybrid mechanism variations with 6 to 10 joints to test the model's adaptability to different numbers of degrees of freedom. However, the core architecture design and final performance evaluation of the model in this paper are both focused on and concentrated on a specific six-DOF hybrid mechanism. The sample construction results are shown in Table 3, covering the number of samples, time window length, number of structural configurations, distribution of joint numbers and statistical characteristics of state variables, verifying the integrity and representativeness of the data organization structure.

TABLE 3. Data set composition and distribution statistics

Dataset Type	Number of Samples	Time Window Length (steps)	Number of Structural Configurations	Joint Count Distribution	Mean (Standard Deviation) of State Variables
Training Set	8500	40	5	6–10	0.01 (0.97)
Validation Set	1062	40	5	6–10	0.00 (1.02)
Test Set	1062	40	5	6–10	0.02 (0.99)

V. RESULTS ANALYSIS

A. FORECAST ACCURACY INDEX PERFORMANCE ANALYSIS

In the trajectory prediction task, to verify the effectiveness of the proposed model in structural modeling and physical consistency constraints, this paper constructs a variety of comparative methods for experimental analysis. The prediction accuracy is evaluated in terms of the end-effector positional error in millimeters. The experimental data is preprocessed in a standardized manner, and the test is carried out under a unified data set and prediction time window. The comparison methods include traditional sequence modeling structures, state space modeling methods, and improved models with different degrees of structure perception. LSTM (Long Short-Term Memory) is the most basic recurrent neural network, which is used to characterize the performance of traditional sequence methods in trajectory prediction tasks. The Mamba model relies on dynamic state modeling to capture long-term dependency information. The standard Transformer model is used to evaluate the generalization ability of the attention mechanism for temporal modeling under unstructured guidance. On this basis, the spatiotemporal attention mechanism is applied to enhance the model's ability to recognize spatial dependencies. Furthermore, the topological structure information is integrated through the DGCN (Dynamic Graph Convolutional Network) to enhance the structural modeling ability. The proposed method adds dynamic consistency constraints on the basis of the aforementioned architecture to improve the physical rationality of prediction and trajectory continuity. The experiment uses two evaluation indicators, RMSE and MAE, to quantitatively compare the numerical accuracy of each prediction model in the hybrid mechanism trajectory output process. The results are shown in Figure

3.

As shown in Figure 3, the RMSE of LSTM is 3.87 and the MAE is 2.95. Its prediction error is mainly affected by its weak long-term dependency modeling ability and lack of structural perception. The RMSE of Mamba is 2.58 and the MAE is 1.92. The state space update mechanism improves the temporal modeling accuracy, but the response to spatial structure is still insufficient. The standard Transformer has an RMSE of 2.74 and an MAE of 2.08. Although the attention mechanism enhances the global modeling capability, it is unstable in processing structural coupling features. After the application of spatiotemporal attention, the RMSE is reduced to 2.18 and the MAE is 1.61. The spatial correlation enhancement mechanism significantly improves the accuracy of multi-modal dynamic capture. After DGCN is integrated, the RMSE is 1.78 and the MAE is 1.32. The graph convolution's ability to model structural changes effectively improves the accuracy of local predictions. On this basis, the method in this paper adds dynamic constraints to reduce the RMSE to 1.42 mm and the MAE to 1.06 mm, which are 63.3% and 64.1% lower than LSTM, respectively. The constraints of dynamic prior to trajectory continuity and physical consistency become the main reason for further reducing the error. Therefore, combining the structure perception mechanism with physical consistency constraints is the key path to achieving high-precision trajectory prediction.

In the trajectory prediction task of the hybrid mechanism, the actual motion path of the end effector in three-dimensional space is often affected by the structural degrees of freedom, joint dynamic disturbances and execution task characteristics. The goal of model training is to obtain the time series dependency between state variables and to accurately fit the mapping relationship between these variables and spatial motion. To verify the ability of the constructed spatiotemporal fusion model to reconstruct motion trends, three comparison graphs of the actual trajectory and the predicted trajectory are constructed on the standard test trajectory, as shown in Figure 4.

The blue curve in Figure 4 represents the actual trajectory of the end effector, which is calculated by forward kinematics from the input state; the red dashed line is the predicted output of the constructed model. The spatial path shows the dynamic evolution of the actuator during the task phase, and the green and black marks correspond to the starting and ending positions of the movement. The two trajectories maintain a high degree of consistency in most sections, reflecting the model's ability to effectively model the dynamic response of the structure. Some local offset areas reveal the sensitivity to local disturbances in time series modeling. This figure provides an intuitive basis for subsequent error quantification and trajectory deviation evaluation.

To analyze the error distribution characteristics of the prediction model on different execution joints, this experiment constructs a six-degree-of-freedom hybrid mechanism with the end effector path as the target, and selects six key joints in the drive chain, corresponding to the front swing arm, auxiliary support rod, main support rod, parallel link, auxiliary actuator arm and terminal rotation joint, numbered 1-6, as the main controlled objects of the prediction model. In order to obtain error distribution data with stable statistical significance, the experiment independently tests the output of each joint position under multiple rounds of simulation tasks. The length of the timing window for each round of acquisition is 60 frames. The spatiotemporal attention fusion Transformer prediction model constructed in this paper is used. After the trajectory prediction is completed, it is compared point by point with the standard sample, and the RMSE and MAE are calculated. The box plot data is generated based on the error values of each time step in the sampling sequence, as shown in Figure 5.

In the RMSE distribution diagram, the median errors of Joint 1 to Joint 3 are 1.42mm, 1.56mm, and 1.66mm, respectively, which are lower than 1.74mm and 1.84mm of Joint 5 and Joint 6, and the maximum error reaches 1.89mm on Joint 6. The error value increases slightly with the increase of joint number, which is mainly caused by the amplification of the downstream joint displacement by the mechanism coupling relationship, and the joint control accuracy is affected by the superposition of the previous stage prediction error. In the MAE distribution, the median values of Joint 1 to Joint 3 are 1.11mm, 1.24mm, and 1.30mm, respectively, which are significantly lower than 1.35mm and 1.42mm of Joint 5 and Joint 6, respectively. The error fluctuation range also shows a similar trend, reflecting the difference in sensitivity of each joint to trajectory disturbance during dynamic response. The model's prediction of the upstream joint is more stable, and the overall error control is still within a reasonable range.

To verify the time response accuracy of the prediction model in different spatial dimensions, a trajectory output experiment of the end effector in three-axis directions is constructed, the test sequence length and input structure parameters are unified to ensure the comparability of the evaluation indicators. By comparing the coordinate differences between the real trajectory and the predicted trajectory at each time step, the fluctuation trend and fitting ability of the observation model in the dynamic prediction process are observed. The three-axis time response of the trajectory is shown in Figure 6.

Figure 6 shows the time-series position change curves of the end effector in the X, Y and Z directions. The solid line is the actual trajectory and the dotted line is the model prediction result. In most time steps, the two sets of curves have the same trend, maintaining good amplitude and phase matching. The predicted curve has a slight deviation in the local time period, indicating that the model has error disturbance in the high-frequency response process in the vertical direction. There is no severe jitter in the overall prediction process, and the sequence error remained stable in each coordinate axis, reflecting the effective decoupling

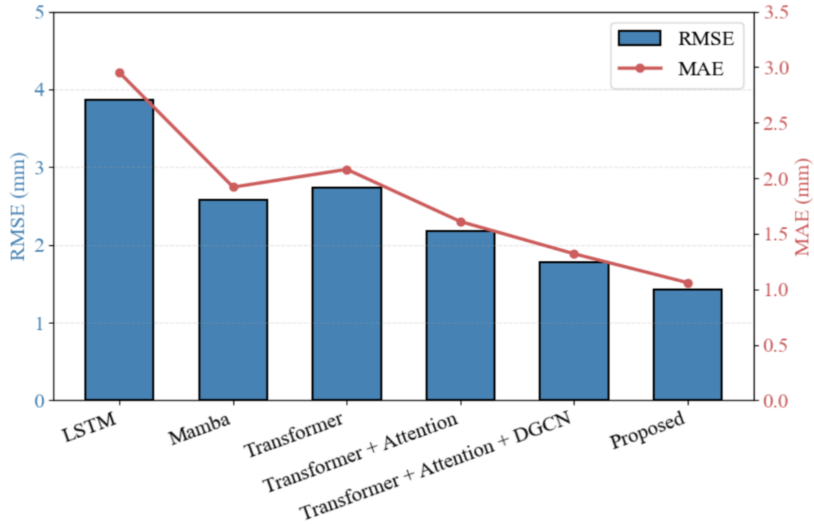

FIGURE 3. Comparison of error indicators of trajectory prediction models

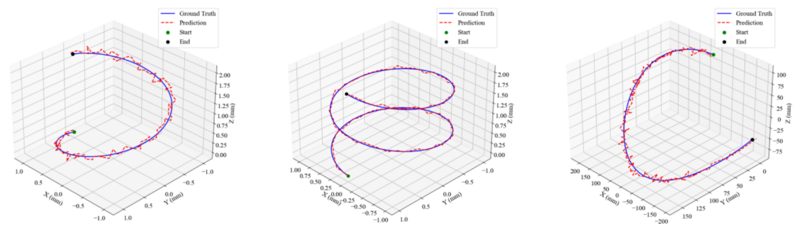

FIGURE 4. Comparison of three-dimensional motion trajectories at the end of the hybrid mechanism

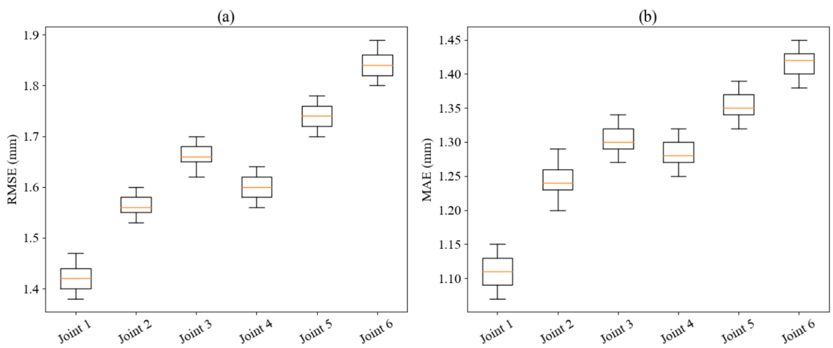

FIGURE 5. Distribution of prediction errors of different joints. Figure 5(a). Comparison of RMSE distribution of each joint; Figure 5(b). Comparison of MAE distribution of each joint Figure 5(c).

ability of the constructed structural modeling module and the spatiotemporal attention mechanism for multi-axis state characteristics.

B. RESPONSE DELAY ANALYSIS

To evaluate the real-time response performance of different model structures in the trajectory correction task, this paper designs a delay test experiment based on the change of input sequence length. The experiment uses a unified hardware environment and reasoning interface to compare the original Transformer structure, the version with spatiotemporal attention module, the version with DGCN module and the complete model with dynamic constraints added in this paper. Each group of models is run under the input sequence conditions of 10 to 100 frames, and the average response delay value during the trajectory correction process is calculated. The impact of the module structure on the prediction efficiency is analyzed by the response performance of different models under various input conditions. The results are shown in Figure 7.

Figure 7 shows that as the input sequence length increases from 10 frames to 100 frames, the response delay of all

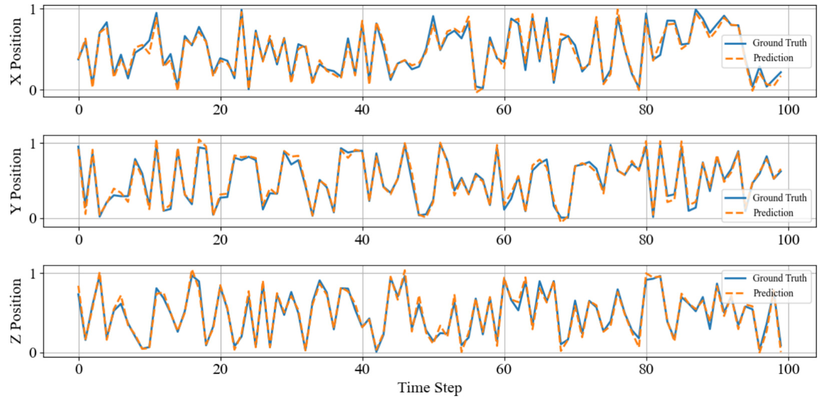

FIGURE 6. Comparison of three-axis position prediction of the end effector

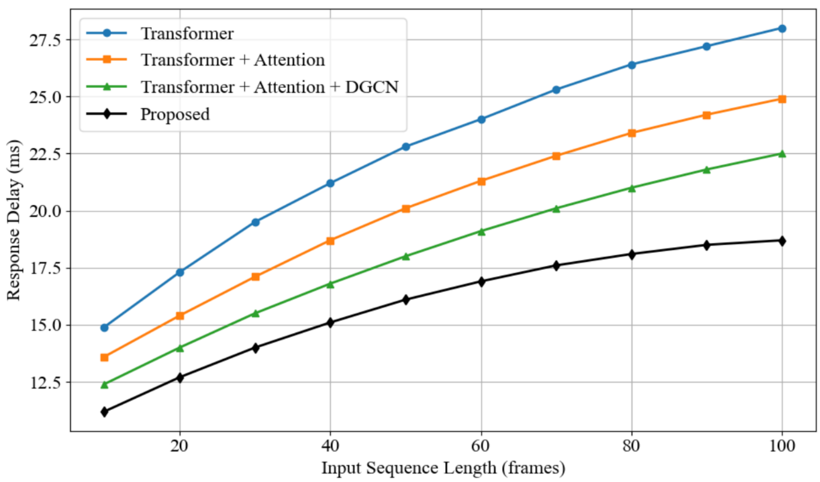

FIGURE 7. Comparison of response delays of different models under varying input sequence lengths

models increases linearly. The response delay of the original Transformer structure increases from 14.9ms to 28.0ms. This structure does not impose any constraints on the feature processing process and has redundant calculations, resulting in a significant decrease in its processing efficiency for long sequences. After applying the spatiotemporal attention module, the model structure focuses on the key feature areas so that the average delay is reduced to 24.9ms at the same length and the calculation process is localized and the reasoning path is shortened. After further adding the DGCN module, the model optimizes the local inter-frame graph structure, reduces unnecessary calculations caused by global self-attention and controls the delay within 22.5ms. After integrating the dynamic constraints, the response delay of the complete model is 18.7ms when the input sequence length is 100 frames. The proposed method maintains a stable and efficient response behavior mainly due to its structured spatiotemporal feature representation and topology-aware modeling, while the physical consistency constraint is applied only during training and does not participate in inference or affect inference efficiency.

C. EVALUATION OF MODEL ADAPTABILITY TO STRUCTURAL DISTURBANCES

We designed a set of trajectory prediction experiments to evaluate the model's generalization and robustness under structural disturbance. The perturbation was simulated by slightly limiting the motion range of a passive joint and fine-tuning a connecting rod's length, thus altering the local motion transmission path without changing overall connectivity—a condition analogous to load disturbance or mechanism looseness in real operation. The process involves: 1) collecting standard data as a reference; 2) using the proposed model (integrating spatiotemporal attention and DGCN) on the perturbed structure; and 3) comparing the true, original, and perturbed prediction trajectories (Figure 8) to intuitively assess the model's ability to adapt to structural uncertainty.

Figure 8 demonstrates the model's adaptability to structural changes. When the hybrid mechanism is stable, the predicted and actual trajectories show high consistency, validating the capture of original dynamic characteristics. After a structural disturbance, although slight point devia-

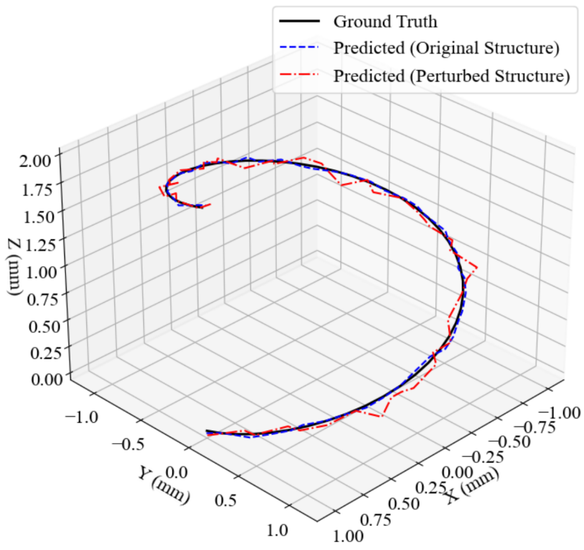

FIGURE 8. Comparison of structural disturbance trajectory prediction

tions occur, the overall trajectory direction remains highly consistent with the true value, reflecting robust modeling capabilities for changing connection paths. This performance is primarily attributed to the Dynamic Graph Convolutional Network (DGCN), which models responses to dynamic changes in mechanism topology, and the spatiotemporal attention mechanism's structural adaptation during feature decoupling. These components significantly improve the system's ability to cope with structural uncertainties under real-world working conditions.

D. PHYSICAL CONSISTENCY VERIFICATION OF TRAJECTORY OUTPUT

In order to verify the performance of the trajectory prediction model in terms of physical consistency, this section analyzes the time continuity of acceleration changes and the stability of the frequency domain. In the experimental setting, the joint acceleration is selected as the key physical quantity, and the model output results before and after the application of dynamic constraints are quantitatively compared. In the time domain, by constructing an acceleration sequence under equal-interval sampling, observe whether there are sudden changes or high-frequency fluctuations in the change trend. In the frequency domain, the frequency distribution in the acceleration signal is extracted based on Fourier transform to reflect the stability of the motion state. Through this process, the comparison results shown in Figure 9 are formed.

Figure 9 shows that the model with dynamic constraints presents a smoother change trend in the acceleration sequence, and the curve changes more gently in the overall fluctuation amplitude and local derivative. The spectrum analysis results show that the high-frequency components of the model are significantly attenuated, the energy is mainly concentrated in the lower frequency band, the frequency response distribution is compact, and no obvious high-frequency surge is shown. In contrast, the unconstrained model has local severe jitters in the acceleration curve, and its spectrum shows multiple high-frequency peaks with prominent energy, indicating that the trajectory has certain defects in time continuity and motion smoothness. This comparison reflects the important role of structural modeling and physical prior injection in dynamic behavior prediction.

E. ANALYSIS OF MODEL COMPLEXITY AND INFERENCE EFFICIENCY

To evaluate the actual performance of the proposed trajectory prediction model in terms of complexity control and reasoning performance, this paper constructs a set of model comparison experiments. The comparison structure includes the basic Transformer model, the enhanced structure that applies the spatiotemporal attention mechanism, the model version that further integrates the dynamic graph convolutional network but does not add physical constraints and the final solution that fully integrates the physical consistency loss. The same test set and hardware platform are used in the experiment to keep the input tensor size consistent with the prediction length to ensure the comparability of inference time and parameter statistics. Based on the standardized test process, the parameter scale and average inference delay per unit sample under each model structure are recorded, and the results are shown in Figure 10.

Figure 10 details the model's computational costs. The base Transformer model has 12M parameters and 10.3 ms latency. Adding the spatiotemporal attention mechanism increases parameters to 15M and latency to 11.8 ms due to nonlinear transformations. Further inclusion of the dynamic graph convolutional network raises parameters to 19M and latency to 15.4 ms, due to dynamic graph updating. The final model contains approximately 20M parameters and exhibits

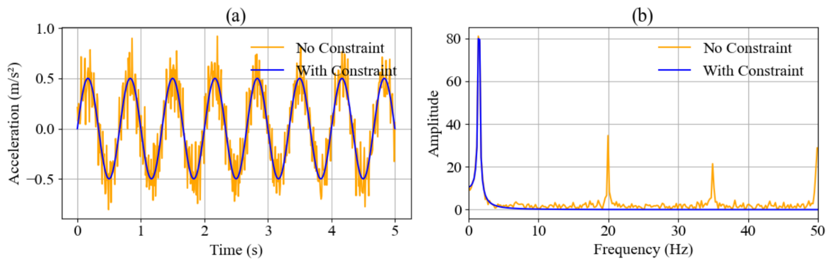

FIGURE 9. Comparison of acceleration time continuity and spectrum response. Figure 9(a). Acceleration time series curve; Figure 9(b). Acceleration spectrum response distribution

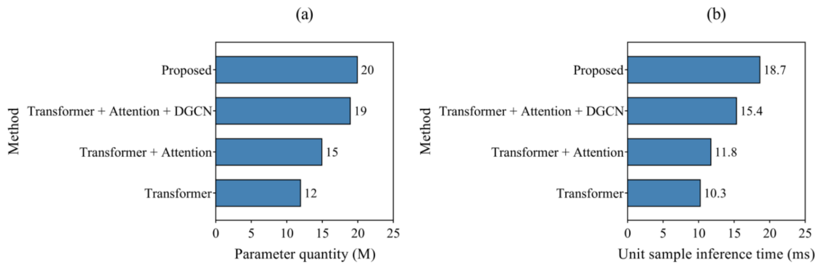

FIGURE 10. Relationship between model complexity and inference efficiency. Figure 10(a). Comparison of model parameters; Figure 10(b). Comparison of unit sample inference time Figure 10(c).

an inference delay of 18.7 ms, which mainly originate from the additional attention layers, dynamic graph convolution operations, and feature fusion processes introduced in the inference network. It should be noted that the Lagrangian-based physical consistency constraint is applied only during the training stage and does not introduce learnable parameters or inference-time computation.

VI. CONCLUSION

This paper proposes a novel Transformer trajectory prediction framework that integrates a spatiotemporal attention mechanism and kinematic constraints to overcome the critical accuracy bottleneck of hybrid mechanisms. This is vital in the industry, where precise, high-speed machinery requires robust modeling for multi-modal feature interaction (like joint force and velocity) and rapid response to structural topology changes, ensuring consistent, defect-free material processing. By constructing a hierarchical spatiotemporal attention module, the decoupled expression of dynamic features such as joint angle, velocity, and torque is realized, and a dynamic graph convolutional network is applied to capture the modulation effect of structural topology changes on the motion path. At the same time, the physical consistency loss function constructed based on the Lagrangian equation is embedded in the training process to ensure the rationality of the predicted trajectory at the dynamic level. The experimental results show that the proposed model achieves an MAE index of 1.06, which is significantly lower than the traditional LSTM and standard Transformer models. In the terminal six-degree-of-freedom joint, the median RMSE of Joint 1 is controlled at 1.42 mm, reflecting the good prediction stability of the upstream structure. In the long sequence prediction task, the response delay under the condition of 100 frames of input is controlled at 18.7 ms, showing excellent real-time performance. The research results confirm that the model maintains stable prediction accuracy and essential real-time response capabilities, even when dealing with complex topological structures and dynamic multi-modal input conditions. This robust performance demonstrates strong potential for both large-scale promotion and significant engineering application value in the high-degree-of-freedom hybrid mechanisms critical to advanced machinery, enabling optimized efficiency and precision in manufacturing.

AUTHOR CONTRIBUTIONS

Haofeng Cao designed, collected and analyzed the data, and drafted the manuscript. Haofeng Cao conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Haofeng Cao participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] X. Fu, D. Zhang, N. Xu, S. Ren, Y. Chu, D. Xiong, et al., "Adaptive Kinematic and Stiffness Control of a Hybrid-Driven Soft Robot for Enhanced Interactions under External Loads," Results in Engineering, vol. 28, pp. 107955-107965, 2025, doi: 10.1016/j.rineng.2025.107955.

[2] A. T. Mathew, D. Feliu-Talegon, A. Alkayas, F. Boyer, and F. Renda, "Reduced order modeling of hybrid soft-rigid robots using global, local, and state-dependent strain parameterization," The International Journal of Robotics Research, vol. 44, no. 1, pp. 129-154, 2025, doi: 10.1177/02783649241262333.

[3] E. Uhlmann, M. N. Fielers, T. Pache, and O. Senyüz, "Stiffness Enhancement by Means of Situational Coupling of Two Collaborative Robots," Journal of Manufacturing and Materials Processing, vol. 10, no. 2, pp. 54-63, 2026, doi: 10.3390/jmmp10020054.

[4] Y. Yang, Z. Chen, C. Guang, Y. Zheng, and C. Lin, "Design and kinematics analysis of a serial-parallel hybrid mechanism used for intraocular surgeries," Journal of Mechanical Engineering, vol. 58, no. 3, pp. 36-44, 2022, doi: 10.3901/JME.2022.03.036.

[5] F. H. Chen, L. P. Zou, and L. Q. Zou, "Research on humanoid robot motion generation and real-time trajectory optimization based on sparse attention mechanism in deep reinforcement learning," Discover Artificial Intelligence, vol. 5, no. 1, pp. 324, 2025, doi: 10.1007/s44163-025-00603-3.

[6] J. Lee and B. K. Cho, "Development of a DDP-based trajectory generation method considering the manipulability measure for 6-DOF collaborative robots," Journal of Computational Design and Engineering, vol. 12, no. 3, pp. 98-114, 2025, doi: 10.1093/jcde/qwae110.

[7] X. B. Chen, H. J. Zhang, H. Yu, J. Liang, and H. Wang, "VNAGT: Variational non-autoregressive graph transformer network for multi-agent trajectory prediction," IEEE Transactions on Vehicular Technology, vol. 72, no. 10, pp. 12540-12552, 2023, doi: 10.1109/TVT.2023.3273230.

[8] B. Mei, Xie Fg, X. J. Liu, Z. H. Xie, and H. C. Zhao, "A mobile hybrid robot and its accuracy issue in machining of large-scale structures," IEEE/ASME Transactions on Mechatronics, vol. 29, no. 1, pp. 347-357, 2023, doi: 10.1109/TMECH.2023.3277459.

[9] G. G. Cui, B. Li, W. Tian, W. H. Liao, and W. Zhao, "Dynamic modeling and vibration prediction of an industrial robot in manufacturing," Applied Mathematical Modelling, vol. 105, pp. 114-136, 2022, doi: 10.1016/j.apm.2021.12.031.

[10] L. H. Li, X. H. Wang, D. F. Yang, Y. F. Ju, Z. G. Zhang, and J. Lian, "Real-time heterogeneous road-agents trajectory prediction using hierarchical convolutional networks and multi-task learning," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 2, pp. 4055-4069, 2023, doi: 10.1109/TIV.2023.3275164.

[11] H. Y. Chen, P. H. Huang, and L. C. Fu, "Social crowd navigation of a mobile robot based on human trajectory prediction and hybrid sensing," Autonomous Robots, vol. 47, no. 4, pp. 339-351, 2023, doi: 10.1007/s10514-023-10103-x.

[12] A. Ahmadi, A. K. Eigoli, and A. Taghvaeipour, "Adaptive backstepping controller based on a novel framework for dynamic solution of an ankle rehabilitation spherical parallel robot," Robotica, vol. 42, no. 5, pp. 1568-1596, 2024, doi: 10.1017/S0263574724000390.

[13] J. K. Yan, L. Jin, Z. T. Yuan, and Z. Y. Liu, "RNN for receding horizon control of redundant robot manipulators," IEEE Transactions on Industrial Electronics, vol. 69, no. 2, pp. 1608-1619, 2021, doi: 10.1109/TIE.2021.3062257.

[14] T. M. N. U. Akhund, Z. A. Shaikh, I. D. L. T. Diez, M. Gafar, D. H. Ajabani, and O. Alfarraj, "IoST-enabled robotic arm control and abnormality prediction using minimal flex sensors and gaussian mixture models," IEEE Access, vol. 12, pp. 45265-45278, 2024, doi: 10.1109/ACCESS.2024.3380360.

[15] C. C. Wang and Y. Q. Zhu, "Identification and Machine Learning Prediction of Nonlinear Behavior in a Robotic Arm System," Symmetry, vol. 13, no. 8, pp. 1445, 2021, doi: 10.3390/sym13081445.

[16] W. Bielajewa, M. Tindall, and P. Nithiarasu, "Comparative study of Transformer-and LSTM-based machine learning methods for transient thermal field reconstruction," Computational Thermal Sciences: An International Journal, vol. 16, no. 3, pp. 89-111, 2024, doi: 10.1615/ComputThermalScien.2023049936.

[17] F. C. Yin, Q. Z. Ji, and C. Z. Wang, "Research on machining error prediction and compensation technology for a stone-carving robotic manipulator," The International Journal of Advanced Manufacturing Technology, vol. 115, pp. 1683-1700, 2021, doi: 10.1007/s00170-021-07230-z.

[18] D. Bruder, D. Bombara, and R. J. Wood, "A Koopman-based residual modeling approach for the control of a soft robot arm," The International journal of robotics research, vol. 44, no. 3, pp. 388-406, 2025, doi: 10.1177/02783649241272114.

[19] J. Z. Ding and C. J. Wang, "Accuracy analysis and error compensation for Tricept machine tool under load," Journal of Mechanical Science and Technology, vol. 35, pp. 3591-3600, 2021, doi: 10.1007/s12206-021-0730-x.

[20] S. Y. Peng, Z. H. Cheng, L. X. Che, S. Cui, and Z. J. Jin, "Configuration design and dimensional synthesis of an asymmetry 2R1T parallel mechanism," Robotica, vol. 41, no. 2, pp. 713-734, 2023, doi: 10.1017/S0263574722001540.

[21] Y. P. Gong and J. B. Lou, "Natural frequency prediction of the 3-RPS parallel manipulator using the substructure synthesis technology," Robotica, vol. 41, no. 9, pp. 2850-2860, 2023, doi: 10.1017/S0263574723000723.

[22] A. Antonov, "Parallel–serial robotic manipulators: a review of architectures, applications, and methods of design and analysis," Machines, vol. 12, no. 11, pp. 811-884, 2024, doi: 10.3390/machines12110811.

[23] G. Singh and V. K. Banga, "Combinations of novel hybrid optimization algorithms-based trajectory planning analysis for an industrial robotic manipulators," Journal of Field Robotics, vol. 39, no. 5, pp. 650-674, 2022, doi: 10.1002/rob.22069.

[24] C. Garriz and R. Domingo, "Trajectory optimization in terms of energy and performance of an industrial robot in the manufacturing industry," Sensors, vol. 22, no. 19, pp. 7538, 2022, doi: 10.3390/s22197538.

[25] A. Rout, D. Bbvl, B. B. Biswal, and G. B. Mahanta, "Optimal trajectory planning of industrial robot for improving positional accuracy," Industrial Robot: the International Journal of Robotics Research and Application, vol. 48, no. 1, pp. 71-83, 2021, doi: 10.1108/IR-07-2019-0148.

[26] G. Singh and V. K. Banga, "Kinematics and trajectory planning analysis based on hybrid optimization algorithms for an industrial robotic manipulators," Soft Computing, vol. 26, no. 21, pp. 11339-11372, 2022, doi: 10.1007/s00500-022-07423-y.

[27] M. G. Tamizi, M. Yaghoubi, and H. Najjaran, "A review of recent trend in motion planning of industrial robots," International Journal of Intelligent Robotics and Applications, vol. 7, no. 2, pp. 253-274, 2023, doi: 10.1007/s41315-023-00274-2.

[28] Y. M. Li, Z. D. Wang, H. Yang, H. Zhang, and Y. K. Wei, "Energy-optimal planning of robot trajectory based on dynamics," Arabian Journal for Science and Engineering, vol. 48, no. 3, pp. 3523-3536, 2023, doi: 10.1007/s13369-022-07185-7.

[29] F. Kordi, S. Mobayen, H. Rezaalikhani, and J. Nikoukar, "Hybrid trajectory tracking control of wheeled mobile robots using predictive kinematic control and dynamic robust control," Asian Journal of Control, vol. 27, no. 3, pp. 1203-1223, 2025, doi: 10.1002/asjc.3503.

[30] R. Xu, J. Li, S. Y. Zhang, L. Li, H. L. Li, G. Y. Ren, et al., "Interactive trajectory prediction for autonomous driving based on Transformer," Mechanical Sciences, vol. 16, no. 1, pp. 87-97, 2025, doi: 10.5194/ms-16-87-2025.

[31] P. Jiang, Z. X. Wang, X. B. Li, X. V. Wang, B. D. Yang, and J. J. Zheng, "Energy consumption prediction and optimization of industrial robots based on LSTM," Journal of Manufacturing Systems, vol. 70, pp. 137-148, 2023, doi: 10.1016/j.jmsy.2023.07.009.

[32] F. Wang, Z. J. Wu, and T. T. Bao, "Time-jerk optimal trajectory planning of industrial robots based on a hybrid WOA-GA algorithm," Processes, vol. 10, no. 5, pp. 1014-1030, 2022, doi: 10.3390/pr10051014.

[33] Y. J. Jiao, M. Z. Miao, Z. S. Yin, C. Y. Lei, X. Zhu, and X. B. Zhao, "A hierarchical hybrid learning framework for multi-agent trajectory prediction," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 8, pp. 10344-10354, 2024, doi: 10.1109/TITS.2024.3357479.

[34] W. H. Zhao, W. Wang, and Z. L. Wan, "Vehicle trajectory prediction based on spatio-temporal Transformer feature fusion," Journal on Communication/Tongxin Xuebao, vol. 45, no. 11, pp. 267, 2024, doi: 10.11959/j.issn.1000-436x.2024192.

[35] Y. L. Liu, J. L. Chen, Y. H. Chang, S. L. He, and Z. T. Zhou, "A novel integration framework for degradation-state prediction via transformer model with autonomous optimizing mechanism," Journal of Manufacturing Systems, vol. 64, pp. 288-302, 2022, doi: 10.1016/j.jmsy.2022.07.004.

[36] J. Q. Zhou, Z. Wang, J. X. Liu, X. L. Luo, and M. Y. Chen, "Modeling and Evaluation of Attention Mechanism Neural Network Based on Industrial Time Series Data," Processes, vol. 13, no. 1, pp. 184-205, 2025, doi: 10.3390/pr13010184.