Pattern Recognition 181 (2027) 114708

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

DG-Game: Dual-view contrastive heterogeneous graph learning for football tactical event classification

Wensheng Wang [a,b], Yudi Huang [c], Haoran Zhou [a,b], Shengpeng Wang [d], Shuli Zhu [a,b], Qingyun Sun [c], Xingcheng Fu [d,*]

[a] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China
[b] Key Laboratory of Target Cognition and Application Technology (TCAT), Beijing, China
[c] School of Computer Science and Engineering, Beihang University, Beijing, China
[d] Key Lab of Education Blockchain and Intelligent Technology, Guangxi Normal University, Guilin, China

ARTICLE INFO

Keywords:
Graph learning
Sports competition

ABSTRACT

Graph Neural Networks (GNNs) have shown strong potential in football event classification, but existing methods usually rely on a single graph structure to model pitch interactions, making it difficult to distinguish cooperative formation organization from adversarial defensive pressure. To address this, we propose DG-Game, a dual-view heterogeneous graph learning framework for football tactical event classification. DG-Game decouples football tracking data into a spatial homogeneous graph and a tactical heterogeneous graph, which respectively characterize teammate collaboration and opponent pressure. Furthermore, dynamic state encoding and multi-scale spatial modeling are introduced to enhance event-level tactical representations, while cross-view constraints promote information complementarity between the two views. Extensive experiments on real-world football tracking datasets demonstrate that DG-Game achieves competitive performance, especially in scenarios requiring joint modeling of cooperative relations, adversarial pressure, and dynamic movement cues.

1. Introduction

With the popularization of advanced tracking technologies, the quantitative analysis of football has become an important research direction in the field of sports analytics [1]. Traditional analysis methods mostly rely on basic statistical features, making it difficult to capture the complex dynamic interactions and collaboration patterns among players. In recent years, Graph Neural Networks (GNNs) have been widely applied in football event classification and event prediction due to their outstanding performance in processing non-Euclidean spatial data. By modeling players and the ball as nodes, and the physical interactions between entities as edges, graph models can effectively extract spatial formation features and achieve real-time evaluation of pitch dynamics [2].

As shown in Fig. 1, although existing graph representation learning methods have made significant progress in football analysis [3,4], their basic modeling paradigm is still to transform tracking data into graph structures by treating players and the ball as nodes and constructing edges according to spatial proximity, passing relations, or predefined interaction rules. However, they still face three main problems. (1) In existing works on homogeneous graph modeling [5], edges are typically constructed based on fixed physical distance thresholds. Team formations require both micro-level physical connectivity and macro-level tactical vision. Constrained by fixed physical distances, homogeneous graphs can only capture local interaction information. (2) Existing works tend to construct a single graph structure, building graphs solely based on passing networks or pure spatial proximity. This single-view modeling approach often ignores the heterogeneous impact of opposing defensive pressure on the ball carrier's decision-making in adversarial sports, leading to the loss of task-relevant information [6]. (3) Modeling approaches from a static perspective, based on the static coordinate positions of players or mere spatial proximity, neglect the instantaneous physical properties of on-field entities. The lack of precise encoding for these high-frequency dynamic states makes it difficult to accurately reflect the micro-level sprint intentions and pressing trends of players.

Recent studies in multi-view graph learning indicate that different views or structural components typically exhibit distinct informational contributions [7]. Processing them identically or prematurely compressing them into a single structure may weaken view discrepancies and lead to insufficient utilization of cross-view information. Regarding football tactical event classification, although teammate collaborative

* Corresponding author.
E-mail address: fuxc@gxnu.edu.cn (X. Fu).

https://doi.org/10.1016/j.patcog.2026.114708
Received 3 May 2026; Received in revised form 22 July 2026; Accepted 17 August 2026
Available online 22 August 2026

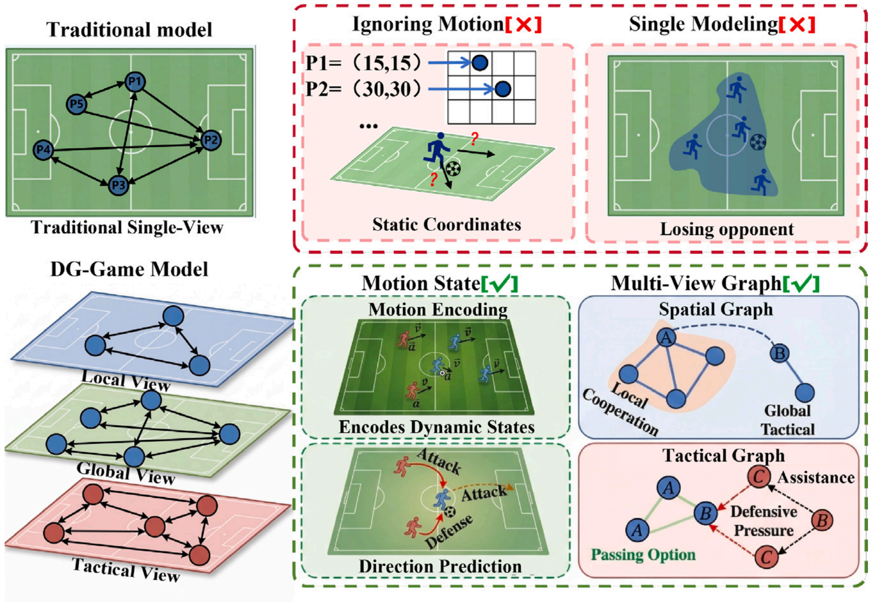

Fig. 1. Illustration of the limitations of conventional single-view football graph modeling and the motivation of DG-Game.

relations and opponent defensive pressure jointly influence the ball carrier's decisions, they exhibit significant differences in tactical semantics. Directly compressing them into a single graph structure may hinder the model from distinguishing between offensive formation support and defensive pressure interference, thereby weakening the critical tactical information required for event classification.

To overcome the representation bottleneck of single-view methods and fully capture the complex cooperative and adversarial interactions on the pitch, this paper explores characterizing pitch dynamics by jointly modeling multi-view graph features. However, effectively applying this multi-view decoupling to dynamic football event prediction tasks still faces significant challenges. (1) Static positions struggle to truly reflect players' movement intentions and team collaboration at a specific moment. How to extract information from discrete event slices and effectively encode players' absolute displacement into instantaneous momentum and movement direction. (2) Entity interactions on the pitch are not simply about physical spatial proximity. During a match, the ball carrier not only needs to evaluate collaboration with friendly players, but must also constantly deal with defensive pressing and passing lane interceptions from opposing players. In dynamic matches, it is difficult to explicitly distinguish between these two entirely different tactical intentions. (3) The spatial modeling of team formations requires both micro-level physical connectivity and macro-level tactical vision. The challenge lies in how to model a spatial graph that contains both local and global interactions. This would allow the homogeneous graph to break through distance limitations, capturing the local short-pass interaction network around the ball carrier while simultaneously discovering the tactical stretching and wide-range transfer intentions of distant players.

To address the aforementioned challenges, this paper proposes DG-Game, a dual-view graph neural network model for football event classification. To capture dynamic match information, we adopt dynamic state encoding. This extracts the instantaneous momentum and movement direction of entities from discrete event slices to construct high-frequency physical features. Next, we address the inability of a single perspective to accurately capture tactical intentions by analyzing physical spatial distances and tactical interaction rules among players. We explicitly decouple and instantiate the pitch topology into a spatial homogeneous graph and a tactical heterogeneous graph. Furthermore, homogeneous formation modeling often lacks a multi-scale vision and is limited by fixed physical distances. To overcome this, our spatial homogeneous graph combines local aggregation with a global attention mechanism. This allows it to simultaneously capture micro-level physical connectivity and macro-level tactical formation stretching. Meanwhile, the tactical heterogeneous graph introduces opposing entities to precisely quantify defensive pressure and passing obstacles, achieving effective feature extraction for both cooperative and adversarial relationships. Subsequently, to fully explore the underlying associations between the two views and extract highly discriminative joint representations, we design a cross-view contrastive pre-training mechanism that constructs multiple self-supervised signals within the latent space. DG-Game optimizes network parameters by dynamically balancing the consistency and specificity between the dual views in the latent space, while also applying uniformity constraints on the feature dimensions. Finally, for the downstream tactical event classification, the model introduces an end-to-end MLP classification head optimized via a tactical-aware class-balanced cross-entropy strategy. This effectively overcomes the masking bias that high-frequency routine actions have on sparse key events, ultimately enabling the precise capture of highly difficult tactical intentions. The main contributions are as follows:

- To more clearly characterize the heterogeneous tactical semantics of coexisting collaboration and confrontation in football matches, we propose a dual-view graph modeling approach for football tactical event classification. This approach represents the offensive collaboration relations and the defensive adversarial pressure as a spatial homogeneous graph and a tactical heterogeneous graph, respectively.
- We design the DG-Game dual-view heterogeneous graph learning framework. It models the formation organization and collaboration relations among teammates through the spatial homogeneous graph, and introduces opponent entities through the tactical heterogeneous graph to quantify defensive pressure, interception risk, and passing lane obstruction.
- We further construct a dynamic state enhanced multi-scale tactical representation mechanism. This mechanism unifies instantaneous momentum, movement direction, local physical connectivity, and global formation structure into the dual-view event representation. It promotes information complementarity between the spatial and tactical views through cross-view constraints.

2. Related work

2.1. Football domain analysis

With the proliferation of computer vision and sensor technology, the data granularity of football matches has evolved from basic post-match statistics to high-frequency tracking data and event data [8,9]. Here, we categorize these analysis methods into three distinct types.

Statistical and traditional methods: Early football analysis primarily relied on manually defined feature engineering and statistical models [10], such as using expected goals models to evaluate shot quality [11], or employing statistical models based on possession zones to quantify a team's control of the field [12]. These methods use logistic regression or tree-based models [13,14] for event classification and prediction. While they offer strong interpretability, they struggle to capture the high dimensional dynamic spatial interactions among players. **Sequence and vision methods:** To process continuous tracking frame data, subsequent research introduced Recurrent Neural Networks (RNNs) and Convolutional Neural Networks [15,16]. Researchers convert two dimensional coordinates of players into top-down heatmaps resembling images, using CNNs to extract spatial features [17]; or they treat the match as a time series, employing model like LSTM [18] to predict passing intentions or match events [19]. However, the core of football lies in the highly dynamic relationships between players and the ball. The chain structure is insufficient for explicitly modeling these complex tactical collaborations (such as multi-player runs and zonal pressing) [5]. **Graph models:** Given that a football match is inherently a dynamic system of multi-agent interactions, graph structures have been gradually introduced into football analysis in recent years [20]. By constructing players and the ball as nodes and their distances or passing networks as edges, researchers have begun using graph representations to capture the dynamic changes in spatial structure [21]. However, most existing graph analysis methods are largely confined to homogeneous physical distance networks, often overlooking the essential differences in tactical pressure between opposing teams.

2.2. Graph learning methods

Graph representation learning has been widely applied across numerous interdisciplinary fields. In molecular generation tasks, graph models are used to extract the underlying topological structures of molecules [22]; In educational data mining, graph networks are employed to construct mapping relationships between exercises and latent knowledge skills [23]. Graph neural networks (such as GCN, GAT) were mainly designed for homogeneous graphs [24], updating local representations by effectively aggregating neighbor node features. To alleviate the problem of label scarcity in real-world scenarios, DGI [25] emerged as a self-supervised graph learning methods based on mutual information maximization and contrastive learning, demonstrating strong potential in unsupervised feature extraction.

However, real world data often consists of heterogeneous graphs containing multiple types of entities and relationships [26,27]. Semi-supervised heterogeneous graph learning methods, such as HAN and HGT [28,29], and self-supervised methods, such as DMGI [30] and CPIM [31], model heterogeneous structures through different relation-aware or multi-view aggregation strategies.

For meta-path based methods, this paradigm essentially transforms a heterogeneous graph into a combination of multiple homogeneous graph to mine the homophily within the graph. HERO [32] bypasses the limitations of traditional meta-path modeling, fully leverages the inherent heterogeneity within graph structures, and captures heterogeneous semantics by aggregating information from diverse node types via a heterogeneous encoder. Additionally, multi-view graph learning methods focus on fusing multiple types of graph structural information more effectively. OMCAL [33] reduces redundant information and noise across different views through adaptive low-rank consensus anchor graph learning, thereby obtaining a higher-quality consensus graph structure. FMVDC [34] fuses spectral embedding matrices rather than directly fusing large-scale similarity matrices. This mitigates the two-stage mismatch problem caused by post-discretization, enhancing clustering performance while maintaining computational efficiency in large-scale scenarios. OMC-ADM [7] introduces an anchor differentiation mechanism to adaptively model the contributions of different anchors. UMCGL [35] learns a universal multi-view consensus graph by jointly modeling the original and generated graphs, effectively capturing cross-view consistency while preserving complementary diversity among different views. DSDGC [36] introduces a dual self-supervised deep graph clustering framework, which enhances graph representation quality through weighted contrastive learning and reliable high-confidence pseudo-label alignment. CTRL [37] focuses on incomplete multi-view multi-label classification, effectively learning compact semantic representations and generating reliable pseudo-labels under missing-view and missing-label settings. These methods demonstrate that different graph structures or views typically contain complementary information. Effective multi-view graph learning should not simply merge different views or process them identically, but should balance cross-view consistency, view-specific diversity, structural complementarity, and reliable supervision.

Although the aforementioned graph representation learning and multi-view structure fusion methods have made progress in general graph modeling, a research gap remains regarding how to specifically model and construct a multi-view graph structure for highly dynamic and strongly adversarial football match scenarios [38].

3. Preliminary

3.1. Problem define

A complete football match can be viewed as a sequential set composed of a series of discrete event slices. Let $\mathcal{S}$ be the set of event slices for a match.

$$\mathcal{S} = \{S_t\}_{t=1}^{T}, \tag{1}$$

where T represents the total number of discrete slices occurring in the match, and S_t represents the tth time slice. Each time slice S_t contains various entity objects participating in the current match event.

We define the entity set $\mathcal{E}$ to contain three different types of core elements: the football, team player, and opposing players:

$$\mathcal{E} = \{b\} \cup \mathcal{U} \cup \mathcal{V}, \tag{2}$$

where b represents the football node, $\mathcal{U} = \{u_1, u_2, \dots, u_M\}$ represents the set of team player nodes, and $\mathcal{V} = \{v_1, v_2, \dots, v_K\}$ represents the set of opposing player nodes. M and K respectively represent the number of team and opposing player on the field.

To overcome the limitations of traditional static coordinate definitions, this model utilizes dynamic state encoding to represent the state of each entity as a continuous feature vector containing instantaneous kinematic attributes. For any entity i in the set $\mathcal{E}$, under a given slice S_t, $\mathbf{x}_{i,t} \in \mathbb{R}^d$ is a continuous feature vector whose state can be represented as:

$$\mathbf{x}_{i,t} = [p_{i,x}, p_{i,y}, c_i, g_i, m_i, \theta_i]^\top, \tag{3}$$

where d represents the dimension of the feature vector. Among the original feature attributes, $p_{i,x}$ and $p_{i,y}$ represent the spatial coordinates of the two-dimensional plane of the entity in the current time slice. c_i is the one-hot indicator variable of the team to which the entity belongs. g_i is the absolute Euclidean distance from the entity to the current attacking target goal. m_i represents the momentum of the entity, calculated through the rate of change of coordinates between consecutive frames. θ_i represents the movement direction of the entity in radians.

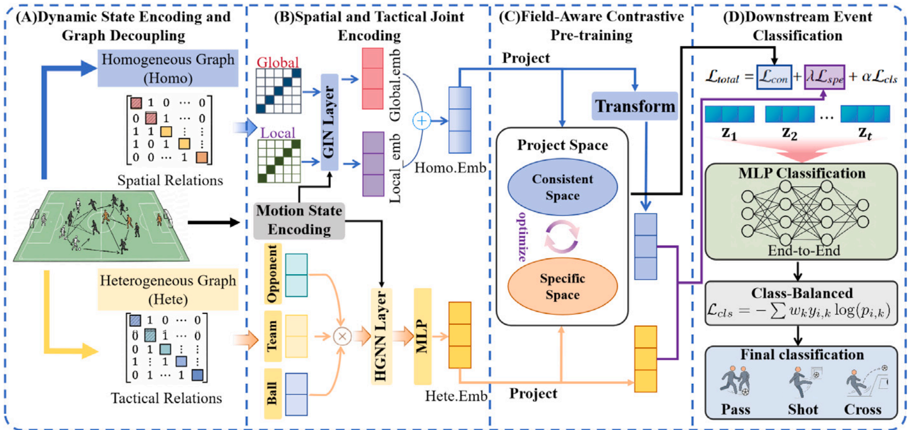

Fig. 2. Illustration of the architecture of DG-Game.

Through the above feature mapping, each discrete football event slice S_t can be transformed into a node feature matrix $\mathbf{X}_t \in \mathbb{R}^{N\times d}$ containing topological and spatiotemporal attributes:

$$\mathbf{X}_t = [\mathbf{x}_{1,t}, \mathbf{x}_{2,t}, \dots, \mathbf{x}_{N,t}]^\top, \tag{4}$$

where N represents the total number of all types of entities in the slice $N = 1 + M + K$. This matrix serves as the underlying data input for the subsequent construction of the homogeneous graph and heterogeneous graph.

3.2. Notations

Based on the node feature matrix $\mathbf{X}_t$, the entity interactions in time slice S_t can be modeled as a topological graph structure. We define the attribute graph $\mathcal{G}_t$ as:

$$\mathcal{G}_t = (\mathcal{E}, \mathbf{A}_t, \mathbf{X}_t), \tag{5}$$

where $\mathbf{A}_t \in \mathbb{R}^{N\times N}$ is the adjacency matrix that quantifies the topological connection relationships and interaction strengths between nodes. For any element $a_{i,j}$ in the matrix $\mathbf{A}_t$, its reflects the association weight between node i and j under specific rules.

$$a_{i,j} = \phi(\mathbf{x}_{i,t}, \mathbf{x}_{j,t}), \tag{6}$$

where $\phi(\cdot,\cdot)$ is the mapping function for calculating the degree of association between entities. When specific edge connection conditions are met between node i and node j, $a_{i,j} > 0$; otherwise, $a_{i,j} = 0$.

Through the formal definition above, isolated entity features are transformed into graph structured data containing spatiotemporal topological information. The subsequent feature aggregation and representation learning will derive specific graph structure representations for the spatial and tactical dimensions, respectively, based on different boundary constraints of the mapping function ϕ.

4. Method

This section proposes DG-Game, a dual-view graph learning framework for football tactical event classification, to capture the complex tactical semantics under the joint effect of collaborative relations and adversarial pressure. Notably, DG-Game aims to learn event-level graph representations, where spatial homogeneous graphs and tactical heterogeneous graphs describe offensive formation collaboration and defensive adversarial interference, respectively. The overall architecture is illustrated in Fig. 2. We first introduce the dynamic state encoding and the dual-view graph construction process, followed by the representation learning methods for the spatial and tactical views. Subsequently, we present the cross-view pre-training objective designed to enhance information complementarity between the two views. Finally, we describe the downstream event classification module and its optimization objective.

4.1. Dynamic state encoding and graph decoupling

This section defines the construction process of the dual-view graph structure. Based on the raw tracking data, we extract a node feature matrix containing the kinematic attributes of entities. Then, relying on this feature matrix, we construct a spatial homogeneous graph by quantifying the physical distances between entities to capture the topological relationships of the friendly formation. Finally, we introduce opposing entity nodes and combine spatial distance with relative movement angles to quantify defensive pressure and passing quality, thereby constructing a multi-relational tactical heterogeneous graph.

Dynamic state encoding. Based on the sequence of original node coordinates, the model quantifies the entity's momentum m_i and movement direction θ_i. Let the duration span corresponding to the time slice S_t be t_Δ, and the spatial displacement coordinate differences of entity i between the start and end frames of the slice be denoted as l_x and l_y, respectively. The average movement speeds of the entity in the horizontal and vertical directions are defined as $v_x = \frac{l_x}{t_\Delta}$ and $v_y = \frac{l_y}{t_\Delta}$, respectively. The momentum m_i is defined as the comprehensive scalar of the entity's movement in the two-dimension plane:

$$m_i = \sqrt{v_x^2 + v_y^2}, \tag{7}$$

where the movement direction θ_i is the angle in radians of the entity's trajectory relative to the coordinate system, define as:

$$\theta_i = \arctan\left(\frac{v_y}{v_x}\right), \tag{8}$$

By integrating entity coordinates, team identifiers, and the distance to the attacking goal, the instantaneous state is initially formed. To effectively capture the continuous temporal dependencies and complex trajectory patterns, the sequence of these instantaneous states across a short temporal window is fed into a Shared Node Motion Encoder instantiated as a Long Short-Term Memory (LSTM) network. The final hidden state is extracted to construct the ultimate node feature matrix $X_t \in \mathbb{R}^{N\times d}$.

Spatial homogeneous graph construction. The spatial homogeneous graph $\mathcal{G}_s$ aims to depict the physical spatial relationships and formation organization among friendly players. We define its set of nodes as: $\mathcal{E}_s = \mathcal{U}$. For any node pair $(i,j) \in \mathcal{E}_s \times \mathcal{E}_s$, let its spatial Euclidean distance be $d_{i,j}$. By introducing a spatial distance threshold τ_s, the element $a_{s,i,j}$ in the spatial adjacency matrix $\mathbf{A}_s$ is defined as the smoothed reciprocal of the distance:

$$a_{s,i,j} = \begin{cases} \frac{1}{d_{i,j}+\epsilon}, & d_{i,j} < \tau_s \\ 0, & d_{i,j} \geq \tau_s, \end{cases} \tag{9}$$

where ϵ represents a small positive smoothing term to prevent division by zero.

Tactical heterogeneous graph construction. The tactical heterogeneous graph $\mathcal{G}_h$ introduces opposing nodes $\mathcal{V}$ to quantify defensive pressure and passing obstacles in an adversarial environment. We define its node set as the complete entity set $\mathcal{E}$. Defensive pressure edges are established between the football node b and the opposing node $v \in \mathcal{V}$. For the team target node $u \in \mathcal{U}$, let the angle in radians α between the vector $\mathbf{l}_v = \mathbf{x}_{v,t} - \mathbf{x}_{b,t}$ and the vector $\mathbf{l}_u = \mathbf{x}_{u,t} - \mathbf{x}_{b,t}$ be defined as:

$$\alpha = \arccos\left(\frac{\mathbf{l}_v^\top \mathbf{l}_u}{\|\mathbf{l}_v\|\|\mathbf{l}_u\|}\right). \tag{10}$$

We define the defensive pressure weight w_d. When $\alpha < \frac{\pi}{2}$, w_d is calculated as:

$$w_d = \frac{1}{d_{b,v} + \epsilon}(1 + \cos\alpha), \tag{11}$$

where $\alpha \geq \frac{\pi}{2}$, this weight reduces to $w_d = \frac{1}{d_{b,v}+\epsilon}$. When w_d exceeds the predefined defensive threshold τ_d, the corresponding element $a_{h,b,v}$ in the heterogeneous adjacency matrix $\mathbf{A}_h$ takes the value of w_d. Passing option edges are established between the football node b and friendly nodes $u \in \mathcal{U}$. For the target node u, we define the cumulative environmental defensive pressure it bears as w_c,

$$w_c = \sum_{v \in \mathcal{V}_u} \frac{1}{d_{u,v} + \epsilon}, \tag{12}$$

where $\mathcal{V}_u$ represents the set of opposing nodes whose spatial distance to node u is less than the interaction threshold. The passing option quality weight w_p is defined as:

$$w_p = \frac{1}{(d_{b,u} + \epsilon)(1 + w_c)}, \tag{13}$$

when the physical distance $d_{b,u}$ is within the effective passing range and w_p exceeds the option threshold τ_p the matrix element $a_{h,b,u}$ takes the value of w_p.

4.2. Spatial and tactical joint encoding

This section explains how to extract high-order event representations from the constructed graph structure through a dual-view graph encoder. The model separately utilizes the topological connectivity of the spatial view and the tactical view to aggregate node features, and maps them to a latent space through deep neural network modules.

Homogeneous spatial encoding. For the homogeneous spatial graph, the encoder combines local physical distances and a global attention mechanism to extract spatial formation features. Breaking through the limitations of fixed physical distances, the encoder further introduces a self-attention mechanism to construct a global adjacency matrix. This enables the model to perceive the tactical stretching and wide-range transfer intentions of distant players.

Given the input node feature matrix $\mathbf{X}_s \in \mathbb{R}^{N_s \times d}$, where N_s represents the total number of nodes in the homogeneous graph, the encoder first calculates the local adjacency matrix $\mathbf{A}_l$. Based on the local graph topology, it utilizes graph isomorphism network layers to aggregate local spatial features,

$$\mathbf{H}_l = f_g(\mathbf{A}_l, \mathbf{X}_s), \tag{14}$$

where f_g represents the aggregation function of the graph isomorphism network, and $\mathbf{H}_l$ represents the aggregated local feature matrix. To capture the team's global tactical formation beyond local distances, the encoder introduces a self-attention mechanism, inspired by the foundational Transformer architecture [39], to construct the global adjacency matrix A_g:

$$\mathbf{A}_g = \sigma\left(\frac{(\mathbf{X}_s\mathbf{W}_q)(\mathbf{X}_s\mathbf{W}_k)^\top}{\sqrt{d_h}}\right), \tag{15}$$

where $\mathbf{W}_q$ and $\mathbf{W}_k$ represent learnable linear projection parameter matrices, d_h represents the hidden layer dimension scalar, and σ represents the softmax function. Based on $\mathbf{A}_g$, another set of graph isomorphism network layers is used to extract the global feature matrix $\mathbf{H}_g$:

$$\mathbf{H}_g = f_g(\mathbf{A}_g, \mathbf{X}_s), \tag{16}$$

After obtaining the local and global features, the encoder compresses the feature dimensions through a pooling mechanism and multi-layer perceptrons (MLPs).

Let $\mathbf{h}_b$ denote the motion-encoded representation of the ball obtained from the shared node motion encoder, and $\mathbf{h}_{l,i}$ and $\mathbf{h}_{g,i}$ represent the feature vectors of the ith team player node under the local and global views, respectively. The calculation process for the local pooled representation $\mathbf{z}_l$ and the global pooled representation $\mathbf{z}_g$ is as follows:

$$\mathbf{z}_l = \left[f_c(\mathbf{h}_b) \parallel f_c\left(\frac{1}{M}\sum_{i=1}^{M}\mathbf{h}_{l,i}\right)\right], \tag{17}$$

$$\mathbf{z}_g = \left[f_c(\mathbf{h}_b) \parallel f_c\left(\frac{1}{M}\sum_{i=1}^{M}\mathbf{h}_{g,i}\right)\right], \tag{18}$$

where f_c represents the dimensionality reduction MLP function, and $\parallel$ represents the vector concatenation operation. The final embedding representation $\mathbf{h}_s$ output by the homogeneous spatial encoder is calculated as follows:

$$\mathbf{h}_s = f_m([\mathbf{z}_l \parallel \mathbf{z}_g]), \tag{19}$$

where f_m represents the feature fusion MLP function, and $\mathbf{h}_s$ represents the final embedding vector output by the homogeneous spatial encoder.

Heterogeneous tactical encoding. The heterogeneous graph encoder aims to integrate the original kinematic information of different types of nodes to quantify tactical defensive pressure. To capture specific adversarial interactions, we inherently perform relation-specific message passing through independent linear projections. Following this relation-specific transformation, the encoder adopts a type-specific feature aggregation strategy. Let $\mathbf{x}_b$ represent the feature vector of the football node, $\mathbf{x}_{u,i}$ represent the feature vector of the ith friendly player, and $\mathbf{x}_{v,j}$ represent the feature vector of the jth opposing player. By performing average pooling operations on the friendly entities and opposing entities separately, and concatenating them with the football node features, the output embedding $\mathbf{h}_h$ of the heterogeneous tactical encoder is obtained:

$$\mathbf{h}_h = \left[\mathbf{x}_b \parallel \frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_{u,i} \parallel \frac{1}{K}\sum_{j=1}^{K}\mathbf{x}_{v,j}\right], \tag{20}$$

where $\mathbf{h}_h$ represents the heterogeneous tactical embedding vector, which contains the global kinematic state distribution of both the offensive and defensive sides.

Feature fusion. After obtaining the independent embeddings from both the spatial and tactical dimensions, the model extracts high order nonlinear interaction features through joint mapping. In practice, the final graph-level representations for both views are obtained by applying a masked mean pooling readout function over all valid entity nodes. The homogeneous spatial embedding $\mathbf{h}_s$ and the heterogeneous tactical embedding $\mathbf{h}_h$ are then concatenated as vectors and input into a deep feed-forward neural network:

$$\mathbf{z} = f_n([\mathbf{h}_s \parallel \mathbf{h}_h]), \tag{21}$$

where f_n represents a deep MLP containing batch normalization operations and ReLU activation functions, and $\mathbf{z}$ represents the final joint representation vector output by the dual-view encoder. This joint representation is subsequently applied to cross-view contrastive pre-training and downstream event classification tasks.

4.3. Field-aware contrastive pre-training

To fully mine the potential associations between the homogeneous graph and heterogeneous graph contrastive learning based on the feature dimensional metrics to construct self-supervised signals.

Given the embedding matrix $\mathbf{H}_s \in \mathbb{R}^{N\times d}$ output by the homogeneous encoder and embedding matrix $\mathbf{H}_h \in \mathbb{R}^{N\times d}$ output by the heterogeneous tactical encoder, the model projects them into a consistency metric space through a non-linear mapping function. We define the consistency projection function $f_g(\cdot)$ to obtain the projection matrices $\mathbf{P}_s = f_g(\mathbf{H}_s)$ and $\mathbf{P}_h = f_g(\mathbf{H}_h)$, where d_g represents the projection space dimension and $\mathbf{P}_s, \mathbf{P}_h \in \mathbb{R}^{N\times d_g}$. To eliminate redundant information among feature dimensions, the model calculates a feature-level uniformity loss $\mathcal{L}_{uni}$. The uniformity loss is formalized as:

$$\mathcal{L}_{uni} = \log \sum \exp(\mathcal{N}(\mathbf{P}_h^\top \mathbf{P}_h)) + \log \sum \exp(\mathcal{N}(\mathbf{P}_s^\top \mathbf{P}_s)), \tag{22}$$

where $\mathcal{N}(\cdot)$ represents the row-level L2-norm normalization operation on the input matrix, $\exp(\cdot)$ represents the exponential mapping of matrix elements, and $\sum$ represents the summation over all matrix elements. This loss function aims to encourage the feature distribution within the same view to tend toward uniformity, pushing apart the distribution distances between different feature dimensions.

To capture the shared essential attributes of football events across different views, the model introduces an invariance loss $\mathcal{L}_{inv}$ in the consistency projection space. This loss aims to minimize the distance between the homogeneous spatial view and the heterogeneous tactical view along the same feature dimensions. The invariance loss is realized by maximizing the diagonal elements of the cross-correlation matrix:

$$\mathcal{L}_{inv} = -\mathrm{Tr}(\mathcal{N}(\mathbf{P}_h^\top \mathbf{P}_s)), \tag{23}$$

where $\mathrm{Tr}(\cdot)$ represents the trace operation that extracts and sums the main diagonal elements of the matrix. While ensuring the consistency of shared information between views, to preserve the unique tactical and spatial information of each view, the model further constructs a specificity loss $\mathcal{L}_{spe}$. It introduces a specificity projection function $f_q(\cdot)$ independent of the consistency projection to obtain the specificity projection matrices $\mathbf{Q}_s = f_q(\mathbf{H}_s)$ and $\mathbf{Q}_h = f_q(\mathbf{H}_h)$. To construct an asymmetric specificity constraint, an additional linear transformation is performed on $\mathbf{Q}_s$ to obtain $\mathbf{Q}'_s = f_p(\mathbf{Q}_s)$. The specificity invariance aims to maintain the orthogonal decoupled state of the dual-view structure in the specificity space.

To prevent the specificity space from collapsing into a trivial solution, the model introduces a non-trivial penalty term. By constraining the distances among the transformed features $\mathbf{Q}'_s$, the original specific features $\mathbf{Q}_s$, and the consistency features $\mathbf{P}_s$, it ensures that the specificity projection provides an effective information gain. The overall specificity loss is formalized follows:

$$\begin{aligned}\mathcal{L}_{spe} = &-\mathrm{Tr}(\mathcal{N}(\mathbf{Q}_s{\prime}^\top \mathbf{Q}_h)) + \eta \left(\mathrm{Tr}(\mathcal{N}(\mathbf{Q}_s{\prime}^\top \mathbf{Q}_s))\right. \\ &\left.+\mathrm{Tr}(\mathcal{N}(\mathbf{Q}_s{\prime}^\top \mathbf{P}_s))\right),\end{aligned} \tag{24}$$

where η represents the hyperparameter controlling the strength of the non-trivial penalty. Finally, the model combines the above graph contrastive losses with auxiliary supervision signals to construct the overall pre-training objective. We define the comprehensive consistency constraint $\mathcal{L}_{con}$ as follows:

$$\mathcal{L}_{con} = \mathcal{L}_{inv} + \gamma \mathcal{L}_{uni}, \tag{25}$$

where γ represents the weight coefficient of the uniformity loss. The model introduces a linear classification head to output predicted logits, and combines them with ground-truth event labels to calculate the cross-entropy classification loss $\mathcal{L}_{cls}$, thereby injecting task-level prior knowledge into the pre-training phase.

The total pre-training loss function $\mathcal{L}_{total}$ is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{con} + \lambda \mathcal{L}_{spe} + \alpha \mathcal{L}_{cls}, \tag{26}$$

where λ represents the weight coefficient of the specificity loss, and α represents the scaling hyperparameter for the auxiliary classification loss. Through an end-to-end gradient backpropagation mechanism, the model continuously optimizes the parameters of the dual-view graph encoder and projectors amid the dynamic interplay of multiple loss functions, ultimately obtaining graph node representations that contain high-order spatiotemporal and tactical coupling relationships.

4.4. Downstream event classification

After completing the cross-view contrastive pre-training, the model extracts the high-order joint representation vector $\mathbf{z}$ of each event slice as the input feature for the downstream classification task. Given a feature matrix $\mathbf{Z} \in \mathbb{R}^{N\times d_z}$ composed of N independent time slices, where d_z represents the vector dimension of the joint representation.

To fully fit the non-linear decision boundaries in the latent space and enable end-to-end optimization, the model adopts a Multilayer Perceptron (MLP) classification head to establish the mapping relationship from features to event labels. For any input joint representation vector $\mathbf{z}_i$, the probability vector output by the classifier is denoted as $\mathbf{p}_i \in \mathbb{R}^C$:

$$\mathbf{p}_i = \mathrm{Softmax}(\mathbf{W}_c \mathbf{z}_i + \mathbf{b}_c), \tag{27}$$

where C represents the total number of football event categories, $\mathbf{W}_c$ and $\mathbf{b}_c$ are learnable weight and bias matrices. The kth element $p_{i,k}$ of the probability vector represents the predicted confidence probability that sample i belongs to category k.

In real football match scenarios, there are significant differences in the decision attributes and the capture difficulty of different tactical events, where passing events dominate, while key tactical events such as shots and tackles belong to the extreme minority classes [40]. Conventional classification optimization relies on standard unweighted losses, which are highly susceptible to being dominated by majority classes in an imbalanced label space, leading to the ignorance of sparse classes.

To overcome the constraint of extreme class imbalance, rather than using manual threshold truncation, the model introduces a Class-Balanced Cross-Entropy (CB-CE) loss strategy to recalibrate the optimization objective. A category-specific weight set $\mathcal{W} = \{w_0, w_1, \ldots, w_{C-1}\}$ is introduced, which dynamically adapts based on the prior probability distribution of the classes in the training data. Let $\pi_k = N_k/N$ denote the empirical frequency of class k, where N_k is the number of samples belonging to class k and N is the total number of training samples. The adaptive penalty weight w_k for class k is formulated as:

$$w_k = \left(\frac{\max_{j\in\{0,\ldots,C-1\}} \pi_j}{\pi_k}\right)^\gamma, \tag{28}$$

where γ is a smoothing hyperparameter that controls the intensity of the weight scaling. Based on this, the downstream classification loss $\mathcal{L}_{cls}$ is defined as:

$$\mathcal{L}_{cls} = -\sum_{k=0}^{C-1} w_k \cdot y_{i,k} \log(p_{i,k}), \tag{29}$$

where $y_{i,k} \in \{0, 1\}$ is the ground-truth label indicator.

The absolute majority class of passes yields the maximal frequency ratio, maintaining a baseline weight (e.g., $w \approx 1$). Conversely, for extreme minority classes, their significantly lower prior probabilities dynamically amplify their loss contribution w_k. The final event prediction label y_i can then be safely and accurately obtained by selecting the category corresponding to the highest calibrated confidence:

$$y_i = \arg\max_k p_{i,k}, \tag{30}$$

where $k \in \{0, \ldots, C-1\}$.

Table 1
Class distribution of the evaluated datasets.

Dataset	Events	Classes	Pass	Shot	Cross
Metrica	2950	2	2882	68	–
Open DFL	6233	3	5833	171	229

5. Experiment

This section will verify the effectiveness of the proposed dual-view graph neural network model in the football event classification task through extensive empirical analysis.

5.1. Datasets and evaluation metrics

Dataset. To comprehensively evaluate the proposed model, we conduct experiments on two real public football match tracking datasets: Metrica and Open DFL [41]. The Metrica dataset is organized in an event-window frame-expanded format. Therefore, we remove duplicated Event_ID records when calculating the event-level class distribution. The Open DFL dataset stores each event as one JSONL record, and graph metadata files are excluded from the event statistics. As shown in Table 1, both datasets exhibit clear long-tailed class distributions. In Metrica, the task is formulated as a binary classification problem with Pass and Shot as the target classes.

The event labels used in this paper are derived from the original dataset annotations. For Metrica, the classification label is obtained from the *Event_Type* field, while *Event_ID*, *Period*, *Event_Start_Frame*, and *Relative_Frame* are used to identify the corresponding event window. Open DFL provides richer contextual annotations. Its labels are obtained from *event_category*, while *event_type* records fine-grained action sources, such as open-play, throw-in, and free-kick events. In addition, qualifier fields provide contextual information including play origin, event outcome, and pressure condition. These annotations enable Open DFL to capture more diverse event sources and tactical contexts than Metrica.

Experimental setup. For the Open DFL dataset, which contains richer tactical descriptions, the raw events are grouped into three discrete classes: Pass, Shot, and Cross. During the graph structure instantiation process, the spatial homogeneous graph is constructed as a topological structure containing 14 nodes, including 14 friendly players. Subsequently, the tactical heterogeneous graph introduces 14 opposing player nodes, expanding into a comprehensive graph structure with 28 nodes to capture adversarial interactions. The feature dimension is uniformly set to 6 dimensions, covering entity coordinates, team identifiers, distance to the goal, momentum, and movement direction. This dataset is extracted from real football match slices (including key events such as shots, passes, ball lost, recoveries, and dribbles). It also contains three different target coordinate categories: Teammate, Opponent, and Ball. The two-dimension spatial coordinates (X, Y) for each category have been recorded in detail. This improves the model's accuracy in detecting, distinguishing, and predicting the spatiotemporal dynamics of these entities in complex football match scenarios. To ensure fair comparison, all methods are trained and evaluated using identical event-level samples, within-match 8:2 splits, raw tracking features, and event labels. We do not impose an identical graph topology on all baselines, as this would either remove their original structural assumptions or transfer the proposed dual-view design to competing methods. Instead, all graph inputs are constructed solely from the same event window: MAGNN uses predefined meta-paths on the heterogeneous event graph, DMGI uses relation-specific adjacency matrices, and TacticAI and C3GNN use the event window as a graph-level sample. No external data, additional event annotations, or DG-Game-specific tactical rules are provided to the baselines. All methods are trained under the same maximum training-epoch budget. Other model-specific hyperparameters follow the recommended settings of the original implementations whenever applicable and are fixed across all match subsets of the same dataset. Our experiments select Macro-F1 (ma-F1) and Micro-F1 (mi-F1) as the evaluation metrics.

5.2. Main experiment analysis

Tables 2 and 3 present a comprehensive performance comparison of all models on the metrica and Open DFL (DFL) datasets, based on Macro-F1 and Micro-F1 metrics.

Baselines model setup. To comprehensively evaluate the performance of DG-Game, we selected six representative categories of graph learning baseline models for comparison:

(1) **Traditional homogeneous graph models**, mainly used to verify the basic ability to aggregate physical spatial topology. Specifically, GCN [42] serves as a classic baseline for semi-supervised homogeneous message aggregation; GIN [43] utilizes sum aggregation and multi-layer perceptrons (MLPs) to maximize discriminative power for evaluating graph isomorphism.

(2) **Heterogeneous graph models**, used to evaluate the performance of modeling complex node relationships and multi-view/meta-paths. This includes DMGI [30], an unsupervised representation learning method that integrates multi-relational embeddings via global mutual information maximization, and MAGNN [44], a heterogeneous graph embedding model that aggregates semantic interactions both within and across predefined meta-paths.

(3) **Domain-specific models**, representing current advanced methods in sports spatio-temporal analysis and complex multi-agent interaction inference. We compare with TacticAI [45], a football tactical AI assistant developed by Google DeepMind that utilizes geometric deep learning for predictive tasks such as receiver identification and shot prediction.

(4) **Imbalanced graph learning models**, representing the latest progress in handling graph long-tail distributions and extreme class imbalance. Specifically, we introduce C3GNN [46], which employs a cluster-guided contrastive learning framework to mitigate long-tail biases. This is used to specifically evaluate the model's performance in capturing sparse, key tactical events.

(5) **Heterogeneous and multiplex graph representation learning methods,** to strengthen the comparison with methods that integrate heterogeneous or multiple graph structures. HeCo [47] employs co-contrastive learning between network-schema and meta-path views to capture local and higher-order heterogeneous graph structures. MGDCR [31] is a multiplex graph representation learning method that reduces intra-graph and inter-graph correlations to suppress noisy information and capture common information across multiple graphs. MGHC [48] further enhances multiplex graph representation learning by restructuring multi-order graph relations to improve homophily and by modeling node-level and class-level consistency across graphs.

(6) **Spatio-temporal multi-agent trajectory modeling methods,** MART [49] employs a hypergraph Transformer to capture pairwise and group-level agent interactions, while MoFlow [50] uses conditional flow matching to model diverse multi-agent motion trajectories. Both methods are adapted to the same event-level classification task and evaluated under the unified experimental protocol.

Results analysis. Tables 2 and 3 present the performance comparison of different models on the Metrica and Open DFL datasets. Overall, DG-Game achieves competitive classification performance on most match subsets, exhibiting more stable performance particularly on the Open DFL dataset. This demonstrates that in scenarios with richer event categories and more complex offensive-defensive interactions, explicitly modeling teammate collaborative relations and opponent defensive pressure provides more effective discriminative information for football tactical event classification.

Table 2
Performance comparison across different games and datasets (ma-F1 metrics). The best results are highlighted in **bold**, and the second-best results are underlined.

Model	Metrica			DFL						
	Game1	Game2	Game3	Game1	Game2	Game3	Game4	Game5	Game6	Game7
GCN	0.6591	0.6407	0.5911	0.4367	0.6239	0.4754	0.6352	0.5245	0.7556	0.7641
GAT	0.6097	0.6295	0.5911	0.5346	0.5519	0.5134	0.5275	0.5302	0.6136	0.7522
GIN	**0.7062**	0.7948	0.6740	0.5738	0.6596	0.4636	0.5085	0.5223	0.6381	0.6138
DMGI	0.5366	0.5008	0.4612	0.4336	0.4058	0.4449	0.4894	0.4654	0.5131	0.5171
MAGNN	0.5023	0.5773	0.5333	0.5343	0.4851	0.5464	0.4091	0.4814	0.5240	0.5899
TacticAI	0.4867	0.4936	0.4956	0.4313	0.4982	0.4271	0.3482	0.3643	0.3940	0.5348
C3GNN	0.6274	0.5672	0.5646	0.6524	0.6594	0.5408	0.4962	0.5234	0.6018	0.6526
HECO	0.5524	0.6452	0.6183	0.5775	0.4950	0.5347	0.5682	0.5237	0.6334	0.7837
MGDCR	0.5524	0.7216	0.6183	0.5062	0.6465	0.5495	0.6076	0.5250	0.6802	0.8196
MGHC	0.5811	**0.8012**	0.6033	0.5749	**0.6737**	0.5072	0.5121	0.5260	0.6383	0.7542
MART	0.5361	0.6621	0.5205	0.5519	0.6318	0.3934	0.4285	0.4508	0.5449	0.6974
MoFlow	0.6274	0.7112	**0.7251**	0.5686	0.4519	0.3582	0.4069	0.4516	0.5122	0.6074
Ours	0.6059	0.7395	0.7167	**0.6813**	**0.6737**	**0.5532**	**0.6832**	**0.5321**	**0.7735**	**0.8293**

Table 3
Performance comparison across different datasets (mi-F1 metrics). The best results are highlighted in **bold**, and the second-best results are underlined.

Model	Metrica			DFL						
	Game1	Game2	Game3	Game1	Game2	Game3	Game4	Game5	Game6	Game7
GCN	0.9152	0.9343	0.9651	0.9104	0.9267	0.8050	**0.9326**	0.8587	0.9162	0.9462
GAT	0.8788	0.9091	0.9651	0.9055	**0.9533**	0.8616	0.8876	0.8967	0.8534	0.9462
GIN	**0.9394**	**0.9798**	0.9694	0.9005	0.9320	0.8805	0.8483	0.8750	0.8691	0.8925
DMGI	0.8242	0.7576	0.7904	0.8557	0.7400	0.7925	0.8315	0.7935	0.7801	0.8172
MAGNN	0.8606	0.8889	0.9258	0.8856	0.8933	0.8050	0.7247	0.8152	0.7487	0.8333
TacticAI	0.8303	0.9747	**0.9825**	0.8507	0.8800	0.8113	0.8764	0.7935	0.8115	0.8656
C3GNN	0.9152	0.8788	0.9520	0.9253	0.9467	0.8365	0.8427	0.8370	0.8429	0.8871
HECO	0.9273	0.9192	0.9738	0.9254	0.9133	0.8616	0.8596	0.8207	0.8639	0.9409
MGDCR	0.9273	0.9646	0.9738	0.9353	0.9400	0.8491	0.8820	0.8478	0.8848	**0.9570**
MGHC	0.9091	0.9747	0.9694	0.9104	**0.9533**	0.8491	0.8427	0.8641	0.8586	0.9355
MART	0.8606	0.9141	0.8646	0.9055	0.9400	0.8050	0.8202	0.7989	0.8272	0.9355
MoFlow	0.9152	0.9394	0.9607	0.9005	0.8467	0.8176	0.8371	0.8261	0.8010	0.9086
Ours	0.9273	0.9596	0.9782	**0.9403**	**0.9533**	**0.9182**	**0.9326**	**0.9076**	**0.9476**	**0.9570**

Nevertheless, DG-Game does not achieve optimal results on all Metrica subsets or across all metrics. For example, GIN achieves higher ma-F1 results on Metrica Game1 and Game2, while MGHC and MoFlow obtain stronger ma-F1 results on Game2 and Game3, respectively. Several baselines also achieve higher mi-F1 scores on individual Metrica subsets. This phenomenon may be related to the task characteristics of Metrica, which contains only Pass and Shot categories with relatively coarse event granularity. In this setting, local spatial structures and short-range interaction cues may already provide sufficient information for binary event discrimination, allowing homogeneous or trajectory-based models to achieve strong results on certain subsets.

In contrast, the Open DFL dataset contains richer tactical event types and more complex match scenarios. The model must simultaneously recognize various events such as passes, shots, and crosses, while processing multiple tactical factors including teammate support, opponent pressure, and passing lane obstruction. In such scenarios, a single spatial topology fails to sufficiently distinguish between collaborative and adversarial relations. Conversely, DG-Game captures event-related tactical semantics more stably by modeling offensive formation organization and defensive pressure interference separately through spatial homogeneous graphs and tactical heterogeneous graphs.

5.3. Ablation study

Contribution analysis of components. Through ablation experiments on key modules, we verified the synergistic effect of each component on model performance.

DG-Game (w/o con) Removes the consistency constraint between the spatial homogeneous view and the tactical heterogeneous view, causing the two views to perform representation learning without explicit shared semantic alignment.

DG-Game (w/o spe) Removes the specificity constraint and retains only the shared representation learning objective between views, thereby weakening the ability to preserve the respective unique information of the spatial and tactical views.

DG-Game (w/o motion) Removes the dynamic state encoding module and uses only static spatial coordinates and basic entity attributes for graph construction and representation learning.

Table 4 presents the ablation results after removing the consistency constraint, specificity constraint, and dynamic state encoding. Overall, the complete DG-Game maintains competitive performance across multiple match subsets, indicating that dynamic state modeling and inter-view constraints can provide effective supplements to dual-view event representations. Meanwhile, we also observe that w/o spe and w/o motion achieve higher results than the complete model on certain Metrica subsets. This indicates that the roles of individual components do not exhibit strictly monotonic gains across all scenarios, but relate to dataset scale, event category complexity, and match scenario differences. For scenarios like Metrica with fewer categories and relatively coarse event granularity, local spatial structures sometimes already provide strong discriminative signals; Consequently, additional dynamic information or view specificity constraints may increase representational complexity and suffer from small-scale data fluctuations. In contrast, in scenarios with more complex offensive-defensive interactions and richer event types, the complete model can more fully fuse dynamic states, shared event semantics, and view-complementary information. Therefore, the ablation results demonstrate that the core components of DG-Game primarily serve multi-source information fusion in complex football tactical events, rather than guaranteeing absolute optimality on all simple subsets.

Table 4
Ablation study on Metrica and DFL datasets. The best results are highlighted in **bold**, and the second-best results are underlined.

Model	Metric	Metrica			DFL						
		Game1	Game2	Game3	Game1	Game2	Game3	Game4	Game5	Game6	Game7
w/o con	ma-F1	0.5759	0.7123	0.6577	0.5976	**0.7857**	0.5852	0.6178	0.5240	0.7314	0.7944
w/o spe	ma-F1	**0.6186**	0.8012	**0.7671**	0.6462	0.6873	**0.5959**	0.5879	0.5085	**0.7850**	0.8174
w/o motion	ma-F1	0.5786	**0.8597**	0.6183	0.6550	0.6388	0.5442	0.6005	**0.5672**	0.6896	0.7526
Ours	ma-F1	0.6059	0.7395	0.7167	**0.6813**	0.6737	0.5532	**0.6832**	0.5321	0.7735	**0.8293**
w/o con	mi-F1	0.9113	0.9435	0.9651	0.9353	**0.9667**	0.9119	0.9101	0.8967	0.9215	0.9462
w/o spe	mi-F1	0.9091	0.9747	**0.9782**	0.9353	0.9533	**0.9182**	0.9101	0.8913	0.9424	0.9516
w/o motion	mi-F1	0.9091	**0.9848**	0.9738	**0.9502**	0.9233	0.9119	**0.9438**	**0.9130**	0.9162	**0.9624**
Ours	mi-F1	**0.9273**	0.9596	**0.9782**	0.9403	0.9533	**0.9182**	0.9326	0.9076	**0.9476**	0.9570

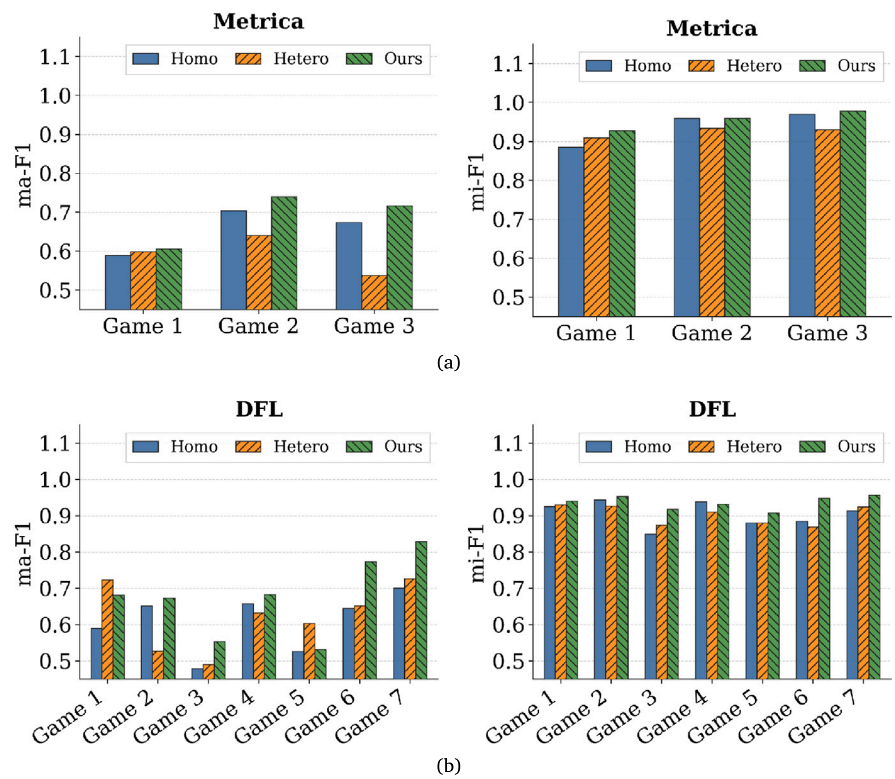

Fig. 3. Analysis of the Micro-F1 (mi-F1) and Macro-F1 (ma-F1) metrics for only the homogeneous graph (Homo), only the heterogeneous graph (Hetero), and our method on the (a) Metrica dataset and (b) DFL dataset.

Spatial graph structure and spatial scale analysis. Figs. 3 and 4 further present the ablation results of the dual-view graph structure and multi-scale spatial modeling.

DG-Game (homo only) Retains only the spatial homogeneous graph and removes the tactical heterogeneous graph, to analyze the impact of offensive formation organization and teammate collaborative relations on event classification.

DG-Game (hetero only) Retains only the tactical heterogeneous graph and removes the independent spatial homogeneous view, to analyze the role of adversarial information such as opponent defensive pressure, interception risk, and passing lane obstruction.

DG-Game (local only) Retains only the local spatial aggregation branch based on physical proximity, to analyze the impact of local short-distance passing and receiving, nearby pressure, and immediate support relations.

DG-Game (global only) Retains only the spatial modeling branch based on global attention, to analyze the impact of macroscopic tactical structures such as distant support, formation stretching, and wide-range transfers.

Overall, the spatial homogeneous graph, the tactical heterogeneous graph, the local spatial branch, and the global spatial branch complement event-level representations from different perspectives. Homo only focuses more on offensive formation organization and teammate collaboration, but lacks opponent defensive pressure information; hetero only can capture defensive interference and passing lane obstruction, but its representation of the overall offensive formation structure is relatively insufficient. Similarly, local only is more suitable for describing short-distance passing and receiving near the ball carrier and local pressure, while global only is more helpful for capturing distant support, weak-side space exploitation, and formation stretching. Because football tactical events are typically influenced simultaneously by local physical connections and global tactical organization, a single view or a single scale may be effective in certain scenarios, but struggles to completely cover the collaborative and adversarial information in complex matches. Consequently, the dual-view graph structure and local–global spatial modeling can provide more complete tactical representations for DG-Game.

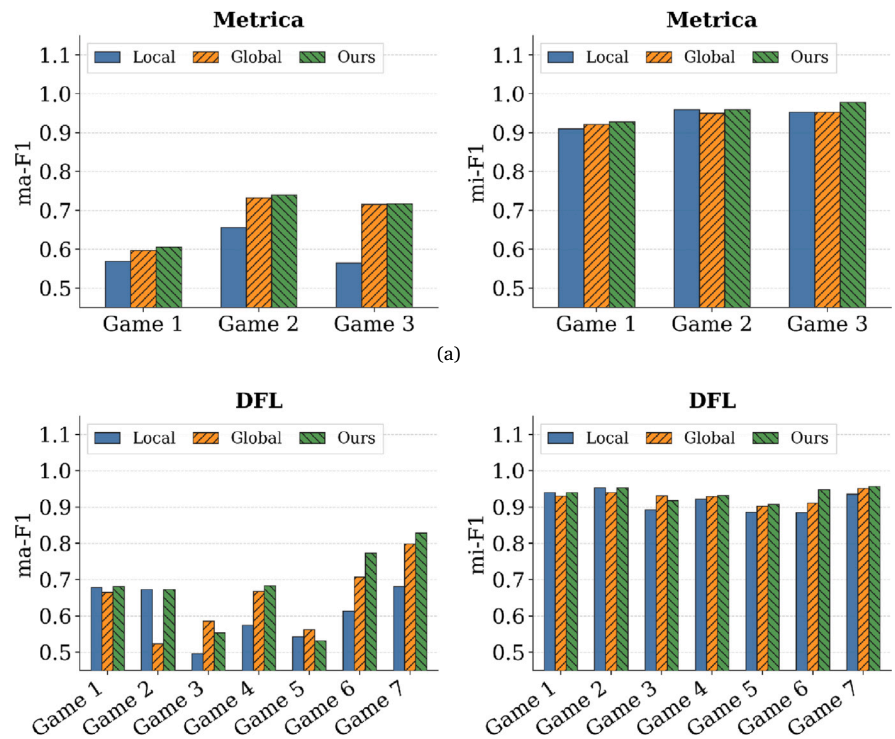

Fig. 4. Analysis of the Micro-F1 (mi-F1) and Macro-F1 (ma-F1) metrics for only the local graph (Local), only the global graph (Global), and our method on the (a) Metrica dataset and (b) DFL dataset.

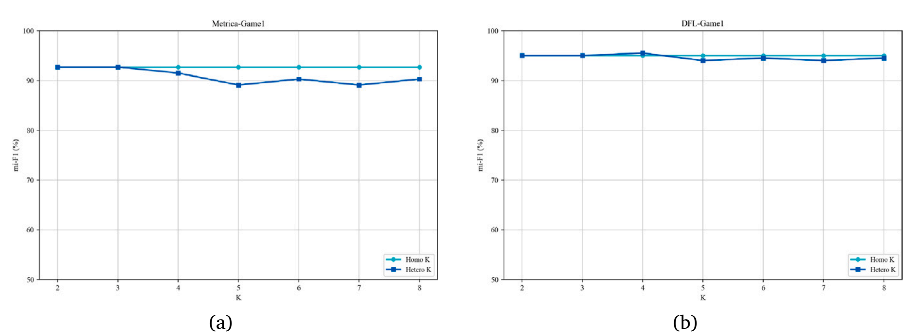

Fig. 5. Sensitivity analysis of the neighborhood size (K), i.e., the number of nearest neighbors retained during graph construction, on Metrica-Game1 and DFL-Game1.

5.4. Hyperparameter sensitivity and convergence analysis

Figs. 5(a) and 5(b) presents representative sensitivity results of the neighborhood size K on Metrica-Game1 and DFL-Game1. The same analysis is conducted on all match subsets. Overall, the performance remains relatively stable as K varies from 2 to 8. The homogeneous-view performance shows only minor changes across different values of K, while the heterogeneous-view performance exhibits moderate fluctuations on Metrica-Game1. These results indicate that the proposed graph construction is robust to the neighborhood size setting within a reasonable range.

To examine the optimization behavior of the joint cross-view objective, Figs. 6(a) and 6(b) presents representative training curves on Metrica-Game1 and DFL-Game1. Although brief fluctuations occur during the early epochs, the total objective decreases consistently and becomes stable in the later training stage on both datasets. The consistency term follows a similar decreasing trend, while the specificity objective gradually approaches a stable value.

6. Conclusion

We propose DG-Game, a dual-view graph neural network model for complex football event classification. It addresses the limitations of traditional static and single-view modeling, which cannot fully capture high-order tactical intentions from dynamic pitch information. Through the spatial homogeneous graph, we primarily characterize formation organization, local connections, and teammate collaborative relations among offensive players; the tactical heterogeneous graph models tactical factors such as defensive pressure, passing obstruction, and adversarial interference by introducing opponent entities. Through this dual-view graph structure design, the model separately captures offensive collaborative information and defensive adversarial information within event-level representations, thereby more accurately describing the heterogeneous interactive relations in football tactical events. Furthermore, the model introduces a dynamic state encoding and a multi-scale spatial modeling mechanism, integrating

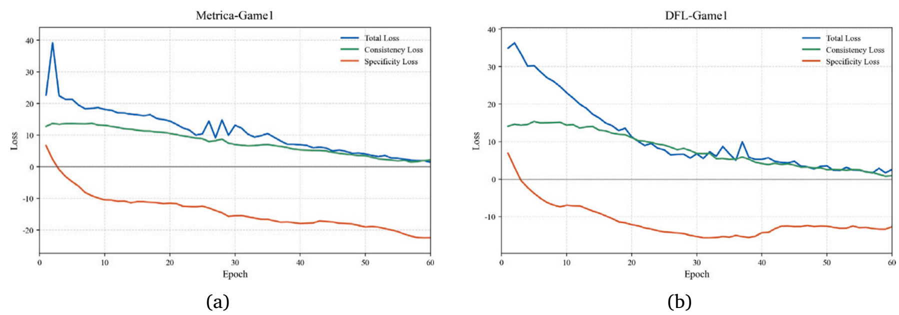

Fig. 6. Representative convergence curves of the total objective, consistency term, and specificity regularization term on Metrica-Game1 and DFL-Game1.

instantaneous player momentum, movement direction, local physical connectivity, and global formation structure into the dual-view event representations to enhance the characterization capability for dynamic match scenarios.

Although this model achieves competitive results in multi-relational graph learning and tactical analysis, there are still some limitations. First, the advantages of DG-Game primarily manifest in tactical scenarios featuring richer event types and more complex offensive-defensive interactions. For binary classification tasks like Metrica, featuring fewer event categories and relatively coarse classification granularity, local spatial structures sometimes already provide strong discriminative information; consequently, certain single-graph structure models may also achieve higher results on specific subsets. Second, constructing the current dual-view topological graphs still relies, to a certain extent, on prior threshold settings such as physical distance and relative angles. Future research will further explore adaptive graph topology inference methods, enabling the model to dynamically learn complex tactical interactive relations in a more data-driven manner. Furthermore, introducing temporal dynamic evolution mechanisms into the current independent discrete time-slice modeling framework to capture long-term tactical evolution and sequential dependencies is also an important research direction. This will further improve the breadth and accuracy of football match predictions in the future.

CRediT authorship contribution statement

Wensheng Wang: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Yudi Huang:** Validation, Software, Data curation. **Haoran Zhou:** Visualization, Validation, Investigation. **Shengpeng Wang:** Writing – review & editing, Methodology, Formal analysis. **Shuli Zhu:** Investigation, Data curation. **Qingyun Sun:** Writing – review & editing, Visualization. **Xingcheng Fu:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Corresponding author is Xingcheng Fu. This work was supported by the Technical Field Fund of the Basic Strengthening Program affiliated with CMC-STC under Grant No. 2025-JCJQ-JJ-0875, the Guangxi Natural Science Foundation under Grant No. 2025JJA170297, the National Natural Science Foundation of China under Grant No. 62462007, and Guangxi Normal University's 2025 Joint Cultivation Project for the National Natural Science Foundation of China under Grant No. 2025PY41. We sincerely thank all the authors for their hard work and contributions.

Data availability

No data was used for the research described in the article.

References

[1] S. Tokuyama, T.C. Greenwell, Examining similarities and differences in consumer motivation for playing and watching soccer, Sport. Mark. Q. 20 (3) (2011) 148–156.
[2] A. Mohandas, M. Ahsan, N. Sarwar, R. Alharthi, M. Latifi, Artificial intelligence for assessing player partnerships in football matches, Discov. Comput. 29 (1) (2026) 104.
[3] F. Majeed, M. Nazir, K. Swart, M. Agus, J. Schneider, Real-time analysis of soccer ball–player interactions using graph convolutional networks for enhanced game insights, Sci. Rep. 15 (1) (2025) 21859.
[4] J. Lee, E. Park, A.P. del Pobil, We know who wins: graph-oriented approaches of passing networks for predictive football match outcomes, J. Big Data 12 (1) (2025) 147.
[5] J. Ochin, G. Devineau, B. Stanciulescu, S. Manitsaris, Game state and spatio-temporal action detection in soccer using graph neural networks and 3d convolutional networks, in: 14th International Conference on Pattern Recognition Applications and Methods, SCITEPRESS-Science and Technology Publications, 2025, pp. 636–646.
[6] W. Ali, S. Vascon, T. Stadelmann, M. Pelillo, Multi-view graph pooling via dominant sets for graph classification, Pattern Recognit. (2025) 112786.
[7] J. Wu, X. Zhang, B. Yang, Y. Fan, Z. Lin, B. Chen, One-pass multiview clustering with anchor differentiation mechanism, Pattern Recognit. 177 (2026) 113286.
[8] M. Manafifard, H. Ebadi, H.A. Moghaddam, A survey on player tracking in soccer videos, Comput. Vis. Image Underst. 159 (2017) 19–46.
[9] L. Pappalardo, P. Cintia, A. Rossi, E. Massucco, P. Ferragina, D. Pedreschi, F. Giannotti, A public data set of spatio-temporal match events in soccer competitions, Sci. Data 6 (1) (2019) 236.
[10] T. Decroos, L. Bransen, J. Van Haaren, J. Davis, Actions speak louder than goals: Valuing player actions in soccer, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 1851–1861.
[11] P. Lucey, D. Oliver, P. Carr, J. Roth, I. Matthews, Assessing team strategy using spatiotemporal data, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 1366–1374.
[12] J. Mead, A. O'Hare, P. McMenemy, Expected goals in football: Improving model performance and demonstrating value, Plos One 18 (4) (2023) e0282295.
[13] C. Herbinet, Predicting Football Results using Machine Learning Techniques (MEng thesis), Imperial College London, 2018.
[14] G. Anzer, P. Bauer, A goal scoring probability model for shots based on synchronized positional and event data in football (soccer), Front. Sport. Act. Living 3 (2021) 624475.
[15] T. Seidl M. Stöckl, D. Marley, P. Power, Making offensive play predictable-using a graph convolutional network to understand defensive performance in soccer, in: Proceedings of the 15th MIT Sloan Sports Analytics Conference, Vol. 2022, 2021, MIT sloan sports analytics conference United States.
[16] G. Capellera, L. Ferraz, A. Rubio, A. Agudo, F. Moreno-Noguer, Transportmer: A holistic approach to trajectory understanding in multi-agent sports, in: Proceedings of the Asian Conference on Computer Vision, 2024, pp. 1652–1670.
[17] B. Aalbers, J. Van Haaren, Distinguishing between roles of football players in play-by-play match event data, in: International Workshop on Machine Learning and Data Mining for Sports Analytics, Springer, 2018, pp. 31–41.
[18] Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: Lstm cells and network architectures, Neural Comput. 31 (7) (2019) 1235–1270.
[19] J. Fernandez, L. Bornn, Wide open spaces: A statistical technique for measuring space creation in professional soccer, in: Sloan sports analytics conference, Vol. 2018, 2018.

[20] T.T. Drexler, Sports analytics with graph neural networks and graph convolutional networks, 2024.
[21] T. Mendes-Neves, J. Mendes-Moreira, R.J. Rossetti, A data-driven simulator for assessing decision-making in soccer, in: EPIA Conference on Artificial Intelligence, Springer, 2021, pp. 687–698.
[22] G. Liu, J. Xu, T. Luo, M. Jiang, Graph diffusion transformers for multi-conditional molecular generation, Adv. Neural Inf. Process. Syst. 37 (2024) 8065–8092.
[23] X. Fu, S. Wang, Y. Gao, X. Li, C. Li, Q. Sun, D. Yu, Towards llm-empowered knowledge tracing via llm-student hierarchical behavior alignment in hyperbolic space, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 40, 2026, pp. 14747–14755.
[24] S. Zhang, H. Tong, J. Xu, R. Maciejewski, Graph convolutional networks: a comprehensive review, Comput. Soc. Networks 6 (1) (2019) 1–23.
[25] P. Velickovic, W. Fedus, W.L. Hamilton, Y. Bengio P. Liò, R.D. Hjelm, et al., Deep graph infomax, ICLR (Poster) 2 (3) (2019) 4.
[26] J. Lv, K. Song, Q. Ye, G. Tian, Semi-supervised node classification via fine-grained graph auxiliary augmentation learning, Pattern Recognit. 137 (2023) 109301.
[27] R. Zheng, W. Chen, G. Feng, Semi-supervised node classification via adaptive graph smoothing networks, Pattern Recognit. 124 (2022) 108492.
[28] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, P.S. Yu, Heterogeneous graph attention network, in: The World Wide Web Conference, 2019, pp. 2022–2032.
[29] Z. Hu, Y. Dong, K. Wang, Y. Sun, Heterogeneous graph transformer, in: Proceedings of the Web Conference 2020, 2020, pp. 2704–2710.
[30] C. Park, D. Kim, J. Han, H. Yu, Unsupervised attributed multiplex network embedding, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 5371–5378.
[31] Y. Mo, Z. Wu, Y. Chen, X. Shi, H.T. Shen, X. Zhu, Multiplex graph representation learning via common and private information mining, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, 2023, pp. 9217–9225.
[32] Y. Mo, F. Nie, P. Hu, H.T. Shen, Z. Zhang, X. Wang, X. Zhu, Self-supervised heterogeneous graph learning: a homophily and heterogeneity view, in: The Twelfth International Conference on Learning Representations, 2024.
[33] Z. Xue, B. Yang, X. Zhang, F. Wang, Z. Lin, One-step multi-view clustering with adaptive low-rank anchor-graph learning, IEEE Trans. Multimed. (2026).
[34] B. Yang, X. Zhang, Z. Xue, F. Nie, B. Chen, Fast multi-view discrete clustering via spectral embedding fusion, IEEE Trans. Pattern Anal. Mach. Intell. (2025).
[35] S. Du, Z. Cai, Z. Wu, Y. Pi, S. Wang, Umcgl: Universal multi-view consensus graph learning with consistency and diversity, IEEE Trans. Image Process. 33 (2024) 3399–3412.
[36] Q. Wang, B. Zhao, Z. Zhang, Q. Gao, L. Jiao, Dual self-supervised deep graph clustering, IEEE Trans. Multimed. (2026).
[37] Y. Liu, C. Liu, J. Wen, L. Shen, B. Zhang, Y. Xu, Learning compact semantic information and reliable pseudo-labels for incomplete multi-view multi-label classification, IEEE Trans. Pattern Anal. Mach. Intell. (2026).
[38] P. Power, H. Ruiz, X. Wei, P. Lucey, Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 1605–1613.
[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in Neural Information Processing Systems, Vol. 30, 2017.
[40] W. Liu, Z. Wu, Y. Wang, H. Ding, F. Liu, J. Lin, G. Lin, Lcreg: Long-tailed image classification with latent categories based recognition, Pattern Recognit. 145 (2024) 109971.
[41] M. Bassek, R. Rein, H. Weber, D. Memmert, An integrated dataset of spatiotemporal and event data in elite soccer, Sci. Data 12 (1) (2025) 195.
[42] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016, arXiv preprint arXiv:1609.02907.
[43] K. Xu, W. Hu, J. Leskovec, S. Jegelka, How powerful are graph neural networks?, 2018, arXiv preprint arXiv:1810.00826.
[44] X. Fu, J. Zhang, Z. Meng, I. King, Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding, in: Proceedings of the Web Conference 2020, 2020, pp. 2331–2341.
[45] Z. Wang, D. Hennes P. Veličković, L. Prince N. Tomašev, M. Kaisers, Y. Bachrach, R. Elie, L.K. Wenliang, F. Piccinini, et al., Tacticai: an ai assistant for football tactics, Nat. Commun. 15 (1) (2024) 1906.
[46] W. Ju, Z. Mao, S. Yi, Y. Qin, Y. Gu, Z. Xiao, J. Shen, Z. Qiao, M. Zhang, Cluster-guided contrastive class-imbalanced graph classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 39, 2025, pp. 11924–11932.
[47] X. Wang, N. Liu, H. Han, C. Shi, Self-supervised heterogeneous graph neural network with co-contrastive learning, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 1726–1736.
[48] Y. Huang, C. Nie, H. He, Y. Mo, Y. Zhu, G. Wen, X. Zhu, Multiplex graph representation learning with homophily and consistency, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 39, 2025, pp. 11835–11842.
[49] S. Lee, J. Lee, Y. Yu, T. Kim, K. Lee, Mart: Multiscale relational transformer networks for multi-agent trajectory prediction, in: European Conference on Computer Vision, Springer, 2024, pp. 89–107.
[50] Y. Fu, Q. Yan, L. Wang, K. Li, R. Liao, Moflow: One-step flow matching for human trajectory forecasting via implicit maximum likelihood estimation based distillation, in: Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 17282–17293.