IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 15 March 2024, accepted 17 March 2024, date of publication 19 March 2024, date of current version 29 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3379410

RESEARCH ARTICLE

STIRNet: A Spatio-Temporal Network for Air Formation Targets Intention Recognition

CHENHAO ZHANG[1], YAN ZHOU[1], HONGQUAN LI[1], YING XU[1], YISHUAI QIN[2], AND LIANG LEI[1]

[1]Air Force Early Warning Academy, Wuhan 430014, China
[2]95980 PLA Troops, Xiangyang 441000, China

Corresponding author: Chenhao Zhang (zch96@foxmail.com)

This work was supported in part by the Graduate Student Sponsored Projects under Grant JY2022C204 and Grant JY2021C178.

ABSTRACT Air formation is a common style of air combat, which demonstrates a high degree of flexibility and strategic value in complex battlefield environments. The activity state of air formation is the result of the intertwining of time domain and air domain, which requires accurate execution of tactical processes in the time axis and skillful deployment of forces in three-dimensional space. Therefore, air formation target combat intention recognition is a complex and challenging task that requires an in-depth understanding of the dynamically changing behavioral patterns of the formation. To address this problem, this paper proposes the STIRNet (Spatio-Temporal Network for Intention Recognition) model, which abstracts the air formation as a spatial graph structure composed of vehicle nodes and combines its temporal data evolving over time. The model autonomously adjusts its attention to different moments and spatial locations through the spatio-temporal attention mechanism, focusing on the important spatio-temporal features that are crucial for recognizing the combat intention of the air formation; and simultaneously captures and integrates the feature information of the air formation in both the temporal and spatial dimensions through the spatio-temporal convolutional operation, which effectively solves the deficiencies of the traditional methods in dealing with the complex spatio-temporal dependency relationships. The experimental results show that the model proposed in this paper effectively improves the accuracy of the combat intention recognition of air formation targets, which is of great value for command decision-making and air battlefield situation assessment.

INDEX TERMS Battlefield situation awareness, air formation targets, intention recognition, spatio-temporal attention, spatio-temporal convolution, spatio-temporal intention recognition network.

I. INTRODUCTION

With the continuous evolution of air warfare patterns and the rapid development of technological advances, air formation plays a crucial role in modern warfare. This highly synergistic combat mode relies on the close cooperation between various combat units within the formation to respond to different mission requirements and to form a synergistic advantage through the rapid transmission and sharing of information. However, in the highly complex and uncertain air battlefield environment, accurately recognizing the combat intention of air formations is a very challenging task, which is of vital significance for accurately and comprehensively grasping the air battlefield situation, making scientific decisions and realizing coordinated strikes.

In earlier studies of target intention recognition, Yang et al. [1] realized the fuzzy inference of the intention of non-cooperative targets by constructing a fuzzy decision tree with K-means clustering for the problem of difficulty in extracting key causal features existing in non-cooperative single targets. However, this method not only has limited capability when dealing with larger datasets, but also needs to discretize the continuous features of the target when dealing with them, which affects the accuracy and applicability of intention recognition. In order to ensure the accuracy of extracting continuous features and the ability of the model

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

to handle large datasets, Ding and Song [2] used the neural network technique to build the FCN-BiGRU (Fully Convolutional Networks - Bidirectional Gated Recurrent Unit) air target intention recognition model, the parallel computing capability, feature extraction capability and adaptive learning capability of neural networks make the model able to deal with large-scale target data efficiently, in addition, the gating structure in the model can extract the continuity features in the target data. In addition, for the problem of target intention recognition in uncertain and incomplete air combat information environment, Xia et al. [3] firstly used cubic spline interpolation and mean complementation to repair the missing target data, and then on the basis of this, they used GRU (Gated Recurrent Unit) network to predict the future state of the target, and then finally constructed a decision tree to reason about its combat intention.

The above studies are all aimed at the intention recognition problem of single target, with the continuous evolution of combat style, more and more scholars are focusing on intention recognition of formation targets. Deng [4] used the HMM (Hidden Markov Model) to speculate the behavioral intent of the formation target based on its formation changes. The combat intention of the formation target is affected by many factors, and this method can not accurately and comprehensively speculate the real intention by only considering its formation. In addition, the current state of the formation target in the HMM is only related to the state of the previous moment without considering the state of the earlier moment, which is unable to capture the dependence of the formation target's formation features over a long distance. Compared to HMM, DBN (Dynamic Bayesian Network) can more flexibly characterize longer time spans and complex nonlinear time dependencies. Yang et al. [5] constructed a DBN that accurately reflects the effect of single-objective uncertainty on the overall intention. In order to further describe the continuity of formation target activities, Yang et al. [6] by building a dynamic sequential Bayesian network, firstly use DBN to capture the changes of formation target states in the time dimension, and then use sequential Bayesian network to infer the combat intention of formation targets. However, the DBN is weak in processing the time series data of long formation targets. Yue [7] designed the LSTM(Long Short-Term Memory) network to solve this problem. By inputting the time series data of formation targets into the network model, the time dependence of formation target features can be explored. Then DBN is used to reasoned the combat intention of the formation target. Each feature of the formation target has different importance for recognizing its combat intent, in order to learn the importance differences between the features of the formation target, Wang et al. [8] introduced the attention mechanism to build an LSTM-Attention network, which learns the weights of each feature of the formation target through the Attenion layer in order to enhance the learning and utilization of the key information of the formation.

The above formation target intention recognition methods focus on exploring the changes of target data over the time dimension, but ignore the effects of interactions and cooperative behaviors among targets within the air formation and their spatio-temporal correlations. In recent years, the rapid development of deep learning and graph neural network (GNN) technologies has brought new opportunities for solving such problems. In this paper, we propose a STIRNet (Spatio-Temporal Intention Recognition Network) model that can simultaneously process spatial features and time-series information to recognize the combat intention of air formation targets, and the main innovations are as follows.

(1) The spatial attention mechanism enables the model to give different levels of attention to information about the relative positions, speeds, and directions between different targets in the formation, while the temporal attention mechanism allows the model to dynamically analyze the sequence of behaviors at different moments and weight them according to the importance of the data at each time point. With the spatio-temporal attention mechanism, the model is able to adaptively learn and assign weights to key time steps and formation members, thus focusing on time information that is critical to recognizing combat intention and key target information in the formation.

(2) STIRNet is able to consider both time series data and spatial topology of air formation targets. Through the spatial convolution operation, it models the spatial relationships within the air formation to capture the synergistic relationship information among the targets; through the temporal convolution operation, the model can capture the dynamic features in the target trajectory data that change in time and space, and efficiently understand the behavioral patterns and their evolutions among the targets in the formation.

(3) In this paper, we compare and analyze the STIRNet model with existing target intention recognition models to validate the work and the effectiveness of its improvement. In addition, in order to verify the effectiveness of each part of the model, we conducted ablation experiments, and the results proved that the STIRNet model can more fully mine and utilize the key spatio-temporal information in the data, and has superior performance in combat intention recognition of air formation targets.

The rest of the paper is organized as follows. Section II provides an overview of the related work on graph neural networks; Section III mathematically describes the combat intention recognition of air formation targets and analyzes the activity state of the air formation; Section IV constructs STIRNet model; Section V conducts experimental validation of the proposed model and compares with other algorithms; and Section VI summarizes the work of this paper.

II. RELATED WORKS

In recent years, neural networks such as convolutional neural networks [9], recurrent neural networks [10], and Transformer [11] have gained powerful characterization and fitting capabilities by constructing models with a large number of covariates, which can not only represent a variety of complex distributions, but also extract high-level abstract features from the original data, and realize the substantial applications of deep learning in computer vision [12], natural language processing [13] and other fields of substantial application. Although the above deep learning methods excel in processing Euclidean data such as images, text and audio, when dealing with non-Euclidean data such as air formation structure, air battlefield knowledge map and command networks, traditional deep learning methods such as convolutional neural networks, recurrent neural networks and Transformer cannot effectively handle such graph data. This is due to the fact that such graph data has arbitrary dimensions and complex topologies that cannot be reduced to matrices or sequences.

In order to solve the above problems, many scholars have started to study Graph Neural Network (GNN) that can process graph data. Sperduti and Starita [14] used recurrent neural networks to process directed acyclic graphs, which is the earliest research on GNN. Gori et al. [15] formally proposed the concept of GNN in 2005. With the wide application of convolutional neural networks in the image field, how to apply the idea of convolutional operations on graph data has become a problem for scholars to think about. Graph Convolutional Neural Network (GCN) is mainly divided into two categories: spectral convolution and spatial convolution. Bruna et al. [16] first proposed a GCN based on spectral graph theory, which implements the convolution operation by Fourier transform, first using the Laplacian matrix of the graph to obtain the Laplacian operator on the frequency domain, and then realizing the convolution of the graph according to the convolution operation of the Euclidean space on the frequency domain. However, the method has more parameters of the convolution kernel, which leads to a significant increase in the cost of the operation, and can only deal with graph data with smaller scale or lower node dimension. To address this problem, Defferrar et al. [17] approximated the representation of the spectral domain convolution with Chebyshev polynomials in order to construct an approximate filter with a finite number of parameters, thus eliminating the need to explicitly solve the spectral decomposition in order to reduce the computational complexity and memory requirements. In addition, Levie et al. [18] can improve the adaptability of GCNs for processing graph data by replacing Chebyshev polynomials with Cayley polynomials. All of the above spectral domain GCNs process graph data with the same structure, but the graph structures abstracted from different scenarios often differ in practical applications. For example, the internal composition and coordination relationship of air target formations performing different missions will not be exactly the same. To solve this problem, Li et al. [19] proposed Adaptive Graph Convolutional Neural Networks (AGCN), which can adaptively learn different representations for input graphs.

Based on the spatial characteristics of graph data, the spatial graph convolution method explores the representation of adjacent nodes in depth, aiming to achieve a unified and orderly representation of adjacent nodes, which is conducive to convolution operations. Niepert et al. [20] proposed the PATCHY-SAN (Select Assemble Normalize) model, which sequentially represents the graph data by selecting the sequence of nodes, calculating the normalized neighborhood graph, and normalizing the neighborhood network, so that the originally complex and varied graph structure can be adapted to a certain extent to the standard convolutional operations, and thus the convolutional operations can be used to mine deep patterns and relationships in graph data. Atwood and Towsley [21] proposed Diffusion Convolutional Neural Networks (DCNN), which utilizes the concept of diffusion process to define a probability transfer matrix on the graph that describes how information is transferred between nodes, so that node information can be gradually aggregated and propagated throughout the graph structure. The size of the receptive field in the GCN determines how many neighbor nodes the network model can aggregate. However, both PATCHY-SAN and DCNN use a recursive way to find neighbor nodes, which leads to the exponential increase of the receptive field with the linear increase of the network model, and the computational cost increases greatly. To solve this problem, Hamilton et al. [22] proposed GraphSAGE (Graph Sample and Aggregate), which dynamically generates a representation of any node by learning a function that aggregates information about neighbor nodes, rather than pre-calculating the fixed embeddings of all nodes. Specifically, it uses a sampling strategy to obtain some neighbors of each node, and integrates the features of these neighbors through a learnable aggregate function to obtain the embedding vector of the target node, making the model inductive and scalable. In the above method, the characteristics of the neighbor nodes are usually aggregated by averaging or weighting the sum of all neighbor nodes, and this type of method gives equal importance to each neighbor node. However, in the actual graph data, there may be significant differences in the influence of different neighbor nodes on the central node. Veličković et al. [23] introduced an attention mechanism and proposed a Graph Attention Network (GAT), which can dynamically assign the importance weight of its neighbor node information to each node, reflecting the degree of contribution of each neighbor to the current node representation update, so that the network model can capture the complex dependencies between nodes and generate discriminative node embeddings.

In reality, the node and edge features of a graph may change dynamically over time. GCN cannot capture the changes of features in the time dimension, and there are limitations in dealing with such problems. In order to solve the problems of time series change, dynamic evolution and order dependence in graph structure data, Graph Recurrent Network (GRN) is widely used in the field of time series graph data. Stochastic Steady-State Embedding (SSE) [24] is a multi-iteration method for processing time series data, in which a feature is randomly selected according to the current state and added to the embedding space with the increase of the number of iterations, and the data gradually tends to be steady-state distributed with the increase of the number of iterations, and the resulting embedding space can capture the main structural information of the data. In addition, Li et al. [25] introduced the gated structure into GNN and proposed a Gated Graph Sequence Neural Network (GGS-NN) The model not only uses GNN to propagate and aggregate information on the graph structure, so that each node can update its own feature representation based on the information of neighbor nodes, but also realizes the long short-term memory function in the graph structure by introducing the update gate and reset gate mechanism, allowing the model to capture and accumulate node state changes across multiple time steps. Considering the differences in the importance of information at different nodes in the graph, Johnson [26] proposed a Gated Graph Transformer Neural Network (GGT-NN) by introducing an attention mechanism based on GGS-NN, which continues to transmit node features through the gating structure, makes the model more flexible and efficient, and also uses the multi-head attention mechanism to enhance the model's ability to model the complex relationship between nodes.

In the above study, GCNs are good at capturing the static or local dynamic spatial dependence between nodes in the graph structure, and usually do not directly consider the changes in the time dimension, so they cannot effectively process the dynamically changing air formation target graph data. Although GRNs processes time series data through cyclic or iterative mechanism, it is not able to model the topology of formation target and the time dependence of target feature simultaneously when processing the time series graph data of air formation target with complex spatial correlation.

III. PROBLEM ANALYSIS

A. PROBLEM DESCRIPTION

In air combat, in the face of complex and high-intensity combat missions, the combat capability of a single target is limited. In order to enhance their own combat effectiveness and adapt to the increasingly complex battlefield environment and the rapidly changing air battlefield situation, they allocate tasks according to the function of each internal target to achieve efficient synergistic operations within the air formation, forming a three-dimensional combat system integrating detection, attack and defense, as well as a multi-level defense depth and attack level, making it difficult for the enemy to deal with all threats at the same time.

In the course of executing a predetermined combat mission, the state and behavior of an air formation target evolves with the changes in spatial and temporal conditions. Their state characteristics essentially constitute time-series data over time, reflecting the process of dynamic change of formation targets. At the same time, in order to adapt to the changing mission requirements and cope with the complex and changeable battlefield environment, the organization and spatial configuration of the air formation must also be flexible and changeable to achieve the optimal combat effectiveness. Based on the above analysis, the target of air formation shows distinctive dynamic characteristics and a high degree of flexibility in both time and space dimensions, which not only needs to effectively execute the established tasks, but also needs to be able to quickly and accurately respond to the ever-changing situation on the battlefield. Therefore, an in-depth study and precise understanding of the behavioral patterns of air formation targets and their implied intentions are of vital significance to a comprehensive assessment of the air battlefield situation. This further proves that through the in-depth mining of the temporal state information and internal structural relationships of air formation targets, we can effectively reverse-engineer their true combat intention.

The air target formation can be abstracted into a graph, each single target of the formation is the node in the graph, and the coordination relationship within the formation is the edge that connects the nodes. An air formation is denoted as $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ represents a single target set in an air formation, $|V| = n$ represents an air formation composed of n single targets, $E = \{e_1, e_2, \ldots, e_m\}$ represents the set of synergies between targets in an air formation, and $|E| = m$ indicates that an air formation contains m synergies, that is, the air formation G contains m edges. The node matrix of air formation G is denoted as $\boldsymbol{X} \in \mathbb{R}^{n\times f}$, where n represents the number of single targets in the air formation and f represents the characteristic dimension of the target. The adjacency matrix of air formation G is defined as

$$\boldsymbol{A}_{ij} = \begin{cases} 1, & \textit{If there is a connection from node i to node j} \\ 0, & \textit{Others} \end{cases} \tag{1}$$

The characteristics and structure of air formation targets change dynamically over time. Assume that the air formation target is abstractly represented as $G^{(t)} = (V^{(t)}, E^{(t)}, \boldsymbol{X}^{(t)})$ at time t. Assume that the target intent is $I = \{i_1, i_2, \ldots, i_p\}$, where i_p denotes the p-th target intention. Then the relationship between the air formation target graph $G^{(t)}$ and the target intention I is

$$I = g\left(G^{(t)}\right) \tag{2}$$

where $g(\cdot)$ represents the mapping relationship between the air formation target graph and the target intention.

B. ANALYSIS OF AIR FORMATION ACTIVITY STATE

As a tactical structure consisting of two or more air units working closely together in a specific formation, the air

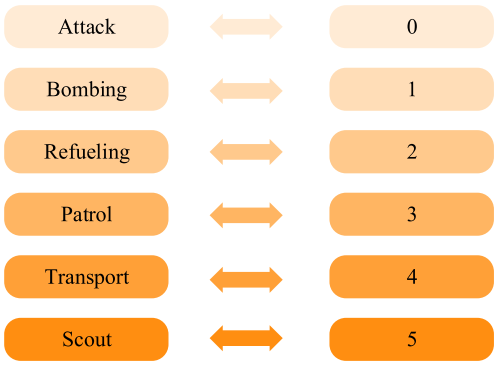


FIGURE 1. Air formation target combat intention space.

formation needs to take into account multiple factors, such as strategic planning, tactical coordination and combat efficiency, when executing a mission, so it shows significant dynamic changes in the time and space dimensions. In this process, each single target in the formation joins the formation sequence and enters the predetermined route according to the specified time window, and may accelerate, decelerate, climb, descend, turn and perform special tactical maneuvers in order to complete the established combat tasks. In order to meet different levels and types of operational requirements, the air formation will organize various types of aircraft according to the actual situation, and carry out formation actions according to different formation configurations. Within the formation, an efficient coordination mechanism is formed among its members to jointly build an intelligent and integrated air combat network.

Air formation performs diverse and complex tasks, which mainly include attack, bombing, refueling, patrol, transport, and scout according to its composition and combat objectives. Attack formation is the core form of organization in air war, which mainly completes the task of striking the air or striking the ground, and usually consists of fighter planes, electronic warfare planes and early-warning planes, etc. The formation type mainly has wedge shape, arrow shape, horizontal shape and vertical shape, etc. Among them, fighter planes carrying various types of precision-guided weapons are responsible for directly carrying out ground strike or air strike missions, and some fighter planes are also responsible for air cover to avoid enemy attacks on formations; electronic warfare planes suppress enemy radars and communication systems by means of electromagnetic interference and electronic countermeasures to create an electromagnetic environment conducive to the penetration of their own formations; and early warning planes provide information support to commanders and combatants through real-time situational awareness of airspace and coordinate the combat operations of the entire formation. Bombing formation is a kind of air group organization method that performs air-to-ground attack to destroy enemy high-value targets with bombers as the core, mainly composed of bombers, fighter planes, electronic warfare planes and early warning planes, etc., and the main formation consists of wedges and trapezoids. Refueling formations are a vital component of modern air warfare, providing fuel to other combat aircraft with tankers to significantly increase their combat radius and duration in the air. In the implementation of aerial refueling mission, refueling aircraft and refueled aircraft to form a specific formation for aerial refueling, in which the recipient aircraft mainly includes fighters, bombers, AWACS (Airborne Warning and Control System), transport aircraft and other types of combat aircraft need to prolong the endurance capacity. In addition, in order to ensure the safety of the refueling process, especially in the enemy threat area, the air refueling formation will also be equipped with a certain number of escort fighters. A patrol formation is an air force unit composed of multiple fighter planes, which conducts continuous surveillance, interception of enemy aircraft, early warning and early strike in designated airspace, and may be equipped with early warning aircraft and electronic warfare aircraft in order to improve situational awareness and jamming capabilities. The transport formation mainly takes large transport aircraft as the core, and other types of support escorts, and lays a solid foundation for winning the war by carrying out tasks such as material transportation, personnel delivery, and rapid deployment of troops during the war or in an emergency. The scout formation is mainly composed of reconnaissance aircraft, fighter aircraft and early-warning aircraft, etc. It carries out scout missions in the airspace or combat area of the enemy by means of the coordination and cooperation between the reconnaissance aircraft and the relevant supporting aircraft, and can obtain the relevant image data, electronic signals and target position information of the enemy. The scout formation is mainly composed of reconnaissance aircraft, fighter aircraft and early-warning aircraft, etc. It carries out scout missions in the airspace or combat area of the enemy by means of the coordination and cooperation between the reconnaissance aircraft and the relevant supporting aircraft, and can obtain the relevant image data, electronic signals and target position information of the enemy. The scout formation provides an indispensable information advantage for its own side by accurately grasping the enemy's situation and responding quickly. In summary, we determine that the intention space of the air formation target includes attack, bombing, refueling, patrol, transport and scout, which contains a total of six kinds of combat intentions, as shown in Figure 1.

Due to the differences in the state of air formation targets when performing different combat missions, their combat intentions can be identified by the state characteristics of formation targets. In this paper, the motion state and sensor state of the target during the mission, as well as its own attributes are comprehensively considered in the construction of the target intention recognition feature space for air formation. There are 10 features in the feature space, including longitude, latitude, azimuth angle, course angle, distance, speed, altitude, sea radar status, air radar state and target type.

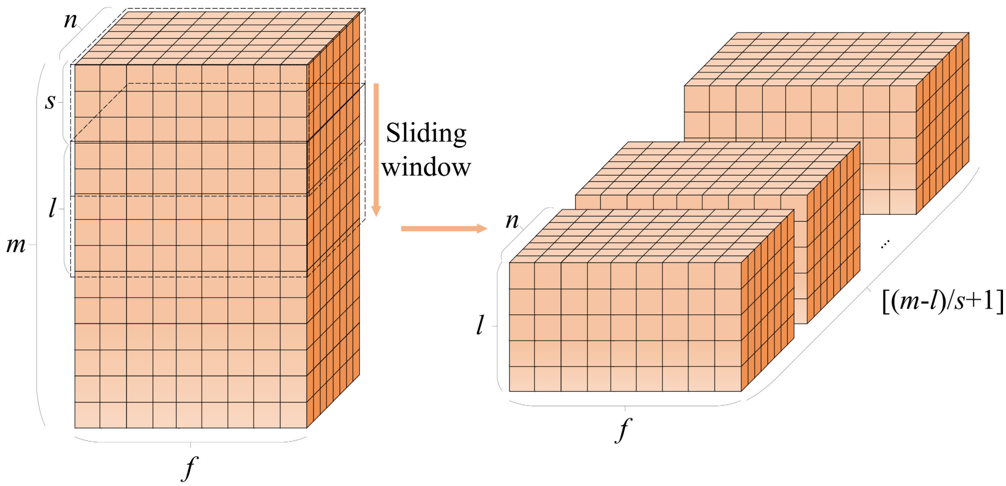


FIGURE 2. Air formation data partitioning.

IV. METHOD

A. DATA PARTITIONING

The situation data of the air formation target has obvious time series characteristics, and the subsequent state of the formation often depends on the previous state. In order to retain the contextual information in the time series when recognizing the combat intention of the air formation and considering the influence of the historical state of the formation, the model needs to use a sliding window to divide the situation data, so as to divide the continuous formation time series data into multiple time segments or subseries with overlapping parts. This method of data partitioning can not only generate a large number of training samples to increase the size of the training set, but also prevent the model from overfitting the features of air formation targets in a specific time period to improve the generalization ability of the model. In addition, dividing the air formation target situation data by a sliding window ensures that each training sample contains time segments of the same length, which is crucial for constructing STIRNet, enabling the network to learn a spatio-temporal feature representation of air formation targets with a fixed window size. The process of sliding window segmentation of air formation target posture data is as follows.

Assuming that an air formation contains n single targets, and each single target has m samples, and creating a sliding window of length l and moving step s, then for each formation data can be divided into $\left[\frac{m-l}{s}+1\right]$ sub-sequence blocks, where $[\cdot]$ denotes the largest integer that does not exceed itself. Then all the target subsequences are pooled to form a four-dimensional data structure of size $\left[\frac{m-l}{s}+1\right] \times n \times l \times f$, where f denotes the dimension of a target sample. The process is shown in Figure 2.

B. DETERMINING THE ADJACENCY MATRIX

In the above, the model divides the situation data of air formation targets into sub-sequences through a sliding window. On this basis, it can be transformed into data of spatio-temporal graph structure. For the air formation, each single target in it can be regarded as a node of the graph, and the characteristics of the node are the state characteristics of the formation target. In order to determine the specific structure of the graph, the relative relationships between nodes need to be clarified, i.e., the adjacency matrix of the graph needs to be determined. It is a core component in the graph structure, which clearly defines the connectivity or interaction between the nodes in the graph, and can represent the relative positions, distances, speed differences and other information between the single targets within the air formation, thus reflecting the spatial structure and dynamic interaction of the formation.

When using STIRNet to identify the combat intention of air formations, the convolution operation enables the capture of feature changes in the spatial and temporal domains, whereas the adjacency matrix enables the model to learn the spatial correlation between different single targets within a formation evolving over time, which is crucial for understanding the combat intention of formations. Under the STIRNet framework, the reasonable construction and use of adjacency matrix to characterize the spatial links and synergistic relationships between single targets of air formations is the basis for accurately recognizing their combat intention. The steps to determine the adjacency matrix are as follows.

Step 1: To determine the large aircraft. According to the formation mission planning, the large aircraft is responsible for providing refueling, information support and other guarantees to other small aircraft types during the mission. Therefore, when determining the structure of the formation graph, there are connections between the large aircrafts and other small aircrafts.

Step 2: Determine the synergy of the rest of the targets based on heading and distance. We divide different heading ranges, and set a distance threshold in each heading range. When the distance between targets is less than the preset threshold, we assume that there is a connection between them in the graph.

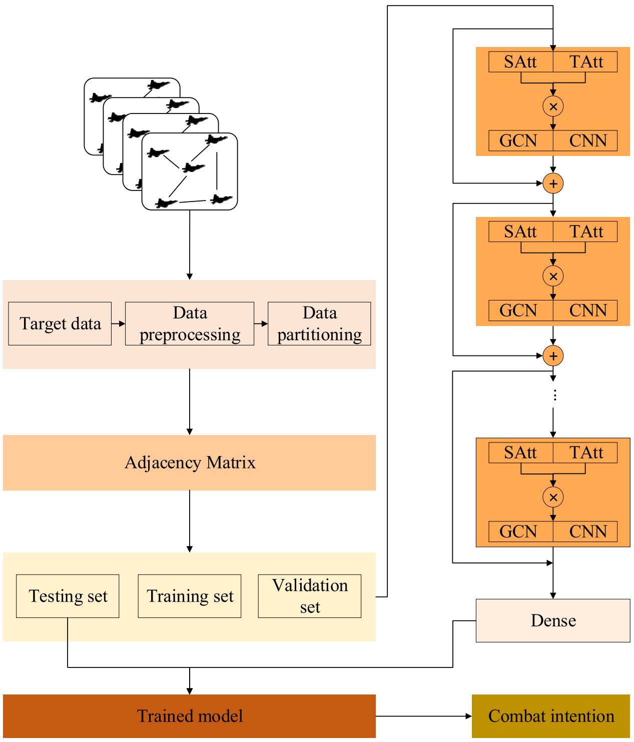


FIGURE 3. **STIRNet model structure.**

Step 3: To create the adjacency matrix of air formation target. According to the connection relationship defined in the above steps, we initialize a matrix of size $n \times n$, where the rows and columns represent the single targets in the formation, and the array elements represent the relationship between the single targets. If there is a connection between two single targets, the value of the corresponding position is 1; otherwise, the value of the corresponding position is 0. In this paper, the elements on the main diagonal of the air formation's adjacency matrix are all 0 and it is a symmetric matrix.

C. STIRNet

STIRNet is able to simultaneously process and analyze the time series data and spatial topology of air formation targets through spatio-temporal attention mechanism and spatio-temporal convolution. The structure is shown in Figure 3. Firstly, data preprocessing is carried out on the air formation target data, including coding, denoising, and normalization, etc.; secondly, the data are divided by sliding window; thirdly, the topology of air formation is determined, i.e., the adjacency matrix of the graph; forthly, the divided target data are divided into the training, validation, and test sets according to the ratio of 6:2:2; fifthly, the training and validation sets are inputted into the spatio-temporal processing module, which extracts the air formation target temporal and spatial features through the spatio-temporal attention mechanism and spatio-temporal convolution, and then the Dense layer is used as the output layer; finally, the test set is inputted into the trained model to recognize the combat intention of the air formation target.

In terms of the computational complexity of the model, the traditional methods are based on sequence models or simple network structures to deal with the spatio-temporal data of air formations, and the computational requirements may increase exponentially with the increase of the number of formation targets, interaction complexity, etc., especially when dealing with complex spatio-temporal dependencies. In contrast, STIRNet models the dynamic behavior of air formations as a graph structure through spatio-temporal graph abstraction, and utilizes GNN for information propagation and feature extraction. Since GNN can effectively process local neighborhood information in parallel, it can reduce the computational complexity compared to the traditional method of global scanning.

In terms of model scalability, traditional methods often struggle to maintain linear growth in performance when dealing with large-scale formations or expanding to scenarios where more targets are involved. STIRNet supports the addition of more targets, and for newly added targets and their spatio-temporal interactions with other targets, it only needs to introduce corresponding new nodes in the network and update the neighboring relationships. This means that the model can effectively handle new input data without significant changes to the overall architecture in the face of changing air formation sizes, thus demonstrating good scalability.

1) SPATIO-TEMPORAL ATTENTION MECHANISM

The spatio-temporal attention mechanism is a computational module that processes the spatio-temporal data of air formation targets in deep learning models, which can be used to understand and model the dependencies in the spatial and temporal dimensions of air formation target data. It is mainly divided into spatial attention mechanism (SAtt) and temporal attention mechanism (TAtt). Air formation flight involves complex mutual cooperation and dynamic changes, and the trajectory, speed, direction, and relative position relationship of each formation member to other members may affect the combat intention of the entire formation. The spatio-temporal attention mechanism assigns different weights to various points in time and synergistic relationships among formation members based on current state and historical information as a way to focus on important points in time and space, highlight key spatio-temporal patterns, and reduce the influence of unimportant or irrelevant factors. For time-series data of air formation targets, the attention mechanism also helps the model to capture and emphasize these long-term temporal dependencies and spatial correlations across multiple nodes of the formation, thus accurately recognizing the combat intention of the formation targets. Combining the spatial and temporal attention mechanisms can jointly optimize the spatial and temporal correlation feature extraction process, enabling the model to more effectively utilize and understand the spatio-temporal structure of the air formation targets, which helps to improve the accuracy and generalization performance of the model.

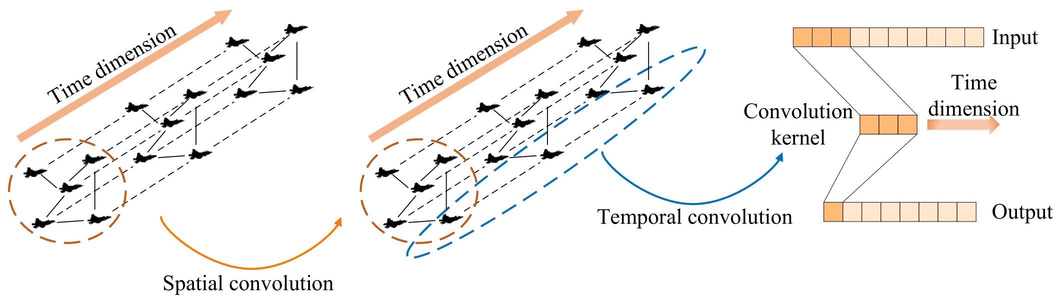


FIGURE 4. Spatio-temporal convolution.

The spatial attention mechanism helps the model learn how to assign weights to the spatial dimensions of the air formation, focusing on different spatial features based on the importance of each node and ignoring unimportant or redundant information [27]. The expression for SAtt is

$$\boldsymbol{SAtt} = \boldsymbol{V}_s \cdot \boldsymbol{\sigma}\left((\boldsymbol{X}\boldsymbol{W}_1)\,\boldsymbol{W}_2\,(\boldsymbol{W}_3\boldsymbol{X})^T + \boldsymbol{b}_s\right) \tag{3}$$

where $\boldsymbol{SAtt}$ denotes the spatial attention matrix; $\boldsymbol{X}$ denotes the input data; $\boldsymbol{V}_s$, $\boldsymbol{W}_1$, $\boldsymbol{W}_2$, $\boldsymbol{W}_3$ and $\boldsymbol{b}_s$ are the parameters to be learned; and σ denotes the Sigmoid activation function.

Normalizing the above equation, we have

$$\boldsymbol{SAtt}'_{i,j} = \frac{\exp\left(\boldsymbol{SAtt}_{i,j}\right)}{\sum_{j=1}^{n}\exp\left(\boldsymbol{SAtt}_{i,j}\right)} \tag{4}$$

where $\boldsymbol{SAtt}_{i,j}$ denotes the correlation strength between the i-th single target and the j-th single target in the air formation.

In view of the changes of the air formation target data in the time dimension, the time attention mechanism can dynamically decide how much attention to be paid to the characteristics of different time steps in the past and the future, so as to capture the short-term and long-term time dependence of the target data [28]. The expression for TAtt is

$$\boldsymbol{TAtt} = \boldsymbol{V}_t \cdot \sigma\left((\boldsymbol{X}\boldsymbol{U}_1)\,\boldsymbol{U}_2\,(\boldsymbol{U}_3\boldsymbol{X})^T + \boldsymbol{b}_t\right) \tag{5}$$

where $\boldsymbol{TAtt}$ is the temporal attention matrix, $\boldsymbol{X}$ is the input data, $\boldsymbol{V}_t$, $\boldsymbol{U}_1$, $\boldsymbol{U}_2$, $\boldsymbol{U}_3$ and $\boldsymbol{b}_t$ are the parameters to be learned, and σ is the Sigmoid activation function.

Normalizing the above equation, we have

$$\boldsymbol{TAtt}'_{i,j} = \frac{\exp\left(\boldsymbol{TAtt}_{i,j}\right)}{\sum_{j=1}^{n}\exp\left(\boldsymbol{TAtt}_{i,j}\right)} \tag{6}$$

where $\boldsymbol{TAtt}_{i,j}$ denotes the correlation strength between the i-th and j-th moments in the air formation.

2) SPATIO-TEMPORAL CONVOLUTION

Spatio-temporal convolution can effectively capture and utilize the spatio-temporal dependencies in the air formation target data, as shown in Figure 4. Spatial convolution can effectively identify and utilize these spatial adjacencies or topologies of air formations to extract features on the airspace. Temporal Convolution is capable of extracting temporal features of formation targets from successive time steps and automatically learns patterns in time-series data to capture local temporal dependencies of such formation target features. By combining the two, the synergistic relationship between different single targets within an air formation within the same time step and the attribute evolution laws of each single target at different moments can be considered simultaneously, which can more accurately fit the complex spatio-temporal behavior of air formation targets.

In air formation graph, each node usually contains feature information of the target and is connected to other nodes through edges to form a complex formation topology. The core idea of GCN is to use the node's own features as well as information of its neighboring nodes to update the node's representation, so as to extract the overall feature representation of the graph [29]. This feature makes GCN able to capture the interdependencies and global contextual information among the nodes within the formation well.

Define the Laplacian matrix for the air formation graph as

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A} \tag{7}$$

where $\boldsymbol{D}$ denotes the degree matrix of the air formation graph, which is a diagonal matrix where the elements on the diagonal are the degree of each vertex in turn, i.e., the number of edges connected to that vertex; $\boldsymbol{A}$ is the adjacency matrix of the air formation graph.

Symmetric normalized Laplacian matrix is obtained by regularizing the Laplacian matrix.

$$\boldsymbol{L}_{sym} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}} \tag{8}$$

Both $\boldsymbol{L}$ and $\boldsymbol{L}_{sym}$ are real symmetric matrices, so they both have n eigenvalues and n eigenvectors that are orthogonal to each other, and both are semi-positive definite matrices. Decompose $\boldsymbol{L}_{sym}$ into $\boldsymbol{L}_{sym} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, where $\boldsymbol{U}$ is a matrix composed of the eigenvectors of $\boldsymbol{L}_{sym}$; $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues of $\boldsymbol{L}_{sym}$. In addition, the eigenvalue range of $\boldsymbol{L}_{sym}$ is [0, 2].

For the air formation graph, it is not possible to give a convolution kernel of a specific shape for the graph because the number of neighboring nodes at each node of the graph is not exactly the same. According to the Fourier transform, the air formation graph is transformed into the spectral domain where the convolution operation is easy to be performed, and after completing the convolution operation, it is inverted in order to return to the spatial domain. Then $\boldsymbol{L}_{sym}$ is multiplied with the node signal $\boldsymbol{x}$ to obtain

$$\boldsymbol{L}_{sym}\boldsymbol{x} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\boldsymbol{x} \tag{9}$$

where $\boldsymbol{U}^T\boldsymbol{x}$ denotes the Fourier transform of the input signal $\boldsymbol{x}$.

Therefore, the convolution operation on the air formation graph only requires computing the symmetric normalized Laplace matrix $\boldsymbol{L}_{sym}$ of the graph and decomposing its eigenvalues and eigenvectors. However, the computational complexity of the feature decomposition of $\boldsymbol{L}_{sym}$ is too high. Spectral convolution on air formation graph can be defined as the multiplication of a nodal signal $\boldsymbol{x}$ with a convolution filter g_θ in the Fourier domain. Define the convolution operation on the air formation graph as

$$g_\theta * \boldsymbol{x} = \boldsymbol{U}g_\theta(\boldsymbol{\Lambda})U^T\boldsymbol{x} \tag{10}$$

$g_\theta(\boldsymbol{\Lambda})$ can be approximated as $g_\theta(\boldsymbol{\Lambda}) \approx \sum_{k=0}^{K} \theta_k \boldsymbol{\Lambda}^k$. However, in practical operations, the problem of vanishing gradient or gradient explosion occurs as the value of K increases. Consider the recursive expression of Chebyshev polynomials as

$$\begin{cases} T_0(x) = 1 \\ T_1(x) = x \\ T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x) \end{cases} \tag{11}$$

The Chebyshev polynomials are defined by $T_k(\cos x) = \cos kx$, which tends to stabilize regardless of the value of K. Therefore, the Chebyshev polynomials do not suffer from vanishing gradient or exploding gradient. The range of the independent variable of the Chebyshev polynomial is $[-1, 1]$, in which case the eigenvalue range of the matrix to be decomposed is D. However, the eigenvalue of $\boldsymbol{L}_{sym}$ can be in the range of $[0, 2]$, so that

$$\hat{\boldsymbol{L}}_{sym} = \boldsymbol{L}_{sym} - \boldsymbol{I} \tag{12}$$

The range of values of the eigenvalues of $\hat{\boldsymbol{L}}_{sym}$ is $[-1, 1]$. Parameterize equation (10) according to the Chebyshev polynomials, that is,

$$\begin{aligned} g_\theta * \boldsymbol{x} &= \boldsymbol{U}g_\theta(\boldsymbol{\Lambda})\boldsymbol{U}^T\boldsymbol{x} \\ &= \boldsymbol{U}\left(\sum_{k=0}^{K}\theta_k T_k(\boldsymbol{\Lambda})\right)\boldsymbol{U}^T\boldsymbol{x} \\ &= \sum_{k=0}^{K}\theta_k \boldsymbol{U}T_k(\boldsymbol{\Lambda})\boldsymbol{U}^T\boldsymbol{x} \\ &= \sum_{k=0}^{K}\theta_k T_k\left(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\right)\boldsymbol{x} \\ &= \sum_{k=0}^{K}\theta_k T_k\left(\hat{\boldsymbol{L}}_{sym}\right)\boldsymbol{x} \end{aligned} \tag{13}$$

Let $K = 1$ in the above equation, equation (13) can be further simplified as

$$\begin{aligned} g_\theta * \boldsymbol{x} &= \theta_0 T_0\left(\hat{\boldsymbol{L}}_{sym}\right)\boldsymbol{x} + \theta_1 T_1\left(\hat{\boldsymbol{L}}_{sym}\right)\boldsymbol{x} \\ &= \theta_0\boldsymbol{x} + \theta_1\hat{\boldsymbol{L}}_{sym}\boldsymbol{x} \\ &= \theta_0\boldsymbol{x} - \theta_1\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{x} \end{aligned} \tag{14}$$

Kipf et al. constrain $\theta_1 = -\theta_0$ [30]. The above equation can be further simplified to

$$g_\theta * \boldsymbol{x} = \theta_0\left(I + \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\right)\boldsymbol{x} \tag{15}$$

Since θ_0 is the parameter to be learned, the above equation can be further simplified to

$$g_\theta * \boldsymbol{x} = \hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{x} \tag{16}$$

where $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}$ and $\hat{\boldsymbol{D}}_{ii} = \sum_j \hat{\boldsymbol{A}}_{ij}$.

To summarize, the information transfer formula between two adjacent graph convolution layers in GCN is

$$\boldsymbol{H}^{(k+1)} = \sigma\left(\hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}^{(k)}\boldsymbol{W}^{(k+1)}\right) \tag{17}$$

where $\boldsymbol{H}$ is the output of the hidden layer, σ is the activation function, and $\boldsymbol{W}$ is the weight parameter that the GCN needs to learn.

Temporal convolution is a key module for processing time-series data of air formation targets, which performs filtering operations on air formation target data in the time dimension through a series of learnable temporal filters to capture formation dynamic features evolving over time. Its expression is

$$\boldsymbol{Z}' = f_r\left(\sum_{\boldsymbol{w}\in M}\boldsymbol{w}\cdot\boldsymbol{Z} + \boldsymbol{b}\right) \tag{18}$$

where $\boldsymbol{Z}'$ denotes the output of the temporal convolution layer; f_r denotes the activation function; $\boldsymbol{w}$ denotes the convolution kernel; M denotes the set of convolution kernels; $\boldsymbol{Z}$ denotes the input of the temporal convolution; and $\boldsymbol{b}$ denotes the bias.

V. EXPERIMENTAL ANALYSIS

A. EXPERIMENTAL DATA AND ENVIRONMENT

The experimental data in this paper comes from the air battlefield posture simulation system. Against the background of airborne multi-service joint operations between the two warring parties, various types of air targets, such as fighter aircraft, transport aircraft, AWACS and reconnaissance aircraft, are simulated to perform the six types of combat missions identified in Section III in an air battle, so as to obtain the state of each target. The target dataset constructed is three-dimensional time-series data, with the first dimension being the number of nodes in each air formation, the second dimension being the number of samples per node, and the

TABLE 1. Results of comparative experiments.

Model	Accuracy (%)	Macro-Precision(%)	Macro-Recall(%)	Macro-F1(%)
FCNN	68	69	69	68
1DCNN	74	73	74	73
RNN	77	77	78	74
LSTM	82	83	82	80
LSTM-Attention	87	87	86	88
STIRNet	**92**	**93**	**93**	**93**

third dimension being the state attributes of each sample. The dataset has a total of 28,263 samples. The data is divided into training set, validation set and test set in the ratio of 6:2:2.

We programm in Python 3.8 on an Ubuntu 20.04 system with the PyTorch 1.11.0 deep learning framework and the Cuda 11.3 parallel computing architecture, the processor is a 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz, the GPU is an RTX 3080, and RAM is 40.0 GB.

B. COMPARATIVE EXPERIMENTS

The proposed STIRNet model is mainly composed of spatio-temporal attention mechanism and spatio-temporal convolution, and the number of spatio-temporal processing modules is set to 2, the order of Chebyshev polynomial is 2, the number of Chebyshev filters and time filters are 32, the Epochs is 60, and the learning rate is 0.0005, cross entropy is used as the loss function and Adam is used as the optimizer.

In order to validate the performance of the STIRNet model in recognizing the combat intention of air formation targets, we design a comparative experiment. The LSTM intention recognition model proposed in [7], the LSTM-Attention intention recognition model proposed in [8], the FCNN (Fully Connected Neural Network) intention recognition model proposed in [31] and [32], and the 1DCNN (One- Dimension Convolutional Neural Network (1DCNN) and RNN (Recurrent Neural Network) are compared with the method proposed in this paper. The results of the comparative experiments are shown in Table 1.

From Table 1, it can be seen that since air formation combat involves highly dynamic and nonlinear spatio-temporal interactions and tactical decision making, and FCNN is a simple structured fully-connected neural network that captures the complex nonlinear relationships of the input data through a multi-layer fully-connected structure, it is not only insensitive to the spatial structure of the air formation targets, but also fails to capture the long-term dependence of the target data in the temporal dimension, and thus has the accuracy is the lowest at 68%. Compared to FCNN, 1DCNN has some advantages in processing time series, and by processing the time series data of air formation targets through the filter, it can better capture the local dependencies in the time series data and the features on the trend of target attribute changes, thus improving the recognition accuracy, and its accuracy rate is 74%. RNN is a neural network designed to process sequential data and is theoretically capable of capturing the time dependence of air formation target data of arbitrary length. However, the original RNN suffers from the problem of gradient vanishing or explosion in the long-term dependency problem, which leads to a mediocre performance for the combat intention recognition containing information dependent on a farther time step, but it has been improved in the short-term dependency so that the accuracy is higher than that of the FCN and the 1DCNN, which is 77%. LSTM is an improved model based on RNN, which effectively solves the long-term dependency problem by introducing a gating mechanism. LSTM makes the model perform well in understanding the temporal evolution of the combat intention of the air formation by retaining the key information from the air formation target data over the past period of time in the network and selectively forgetting or updating the state as needed, which further improves the accuracy rate to 82%. The attention mechanism is combined with LSTM to construct the LSTM-Attention model, in which the attention mechanism allows the model to dynamically focus on the important parts of the air formation target time series data, avoiding the situation of treating all the time steps equally, and helping to improve the sensitivity to the key time nodes of the combat intention and the recognition accuracy, with an accuracy rate of 87%. The proposed STIRNet model combines the spatio-temporal attention mechanism and spatio-temporal convolution operation, which is suitable for processing air formation target time series data with explicit spatial topology and temporal dependence. In air formation combat intention recognition, the model is able to simultaneously consider the spatial relationships between single targets within the formation and their changes over time, and effectively extracts the synergistic relationships within the formation through the operation of the adjacency matrix. As a result, STIRNet demonstrates the strongest performance in understanding and processing complex spatial and temporal dynamic behaviors, thus obtaining the highest combat intention recognition accuracy for air formation targets. Compared with other intention recognition models, the proposed STIRNet model has an average improvement of about 19.41% in Accuracy, 19.09% in Macro-Precision, 18.96% in Macro-Recall, and 19.72% in Macro-F1.

However, it is obvious that STIRNet involves complex spatio-temporal convolution and spatio-temporal attention mechanisms, which are more computationally intensive compared to other models, but it is able to more accurately model formation behavior when dealing with highly correlated spatio-temporal data, and in the long run, the improved model performance may offset the lack of computational efficiency. It should be noted that the operation is only executed on the hardware devices used in this paper, and with the development of hardware devices, we believe that the computational timeliness is not an issue that needs to be considered in particular.

C. ABLATION EXPERIMENTS

The proposed STIRNet model is compared with FCNN, 1DCNN, RNN, LSTM and LSTM-Attention in the above, and the results have fully proved that the model possesses

TABLE 2. Results of ablation experiments.

Model	Accuracy (%)	Macro-Precision (%)	Macro-Recall(%)	Macro-F1(%)
GCN	84	82	82	81
STGCN	87	89	88	88
GAT	89	90	89	89
STIRNet	**92**	**93**	**93**	**93**

superior performance in recognizing the combat intention of air formation targets. In order to analyze and evaluate the impact of each component or module in the model on the overall intention recognition performance, ablation experiments are conducted here and the results are shown in Table 2.

As can be seen from Table 2, the accuracy of the proposed STIRNet model is 9.52%, 5.75% and 3.37% higher than that of GCN, STGCN (Spatio-Temporal Graph Convolutional Network) and GAT (Graph Attention Network), respectively Macro-Precision is 12.20%, 3.37% and 2.22% higher than the above models, Macro-Recall is 12.20%, 4.55% and 3.37% higher than the above models, and Macro-F1 is 12.35%, 3.41% and 2.25% higher than the above models. The GCN model is dedicated to processing non-Euclidean data, and can effectively capture the synergistic relationships and feature propagation among nodes within an air formation. However, for scenarios such as air formations, which contain complex spatio-temporal dynamic characteristics, GCN can only extract the spatial correlation between the internal formations, and does not have the ability to directly deal with the target's time series changes, so it has the lowest intention recognition accuracy. On the basis of GCN, STGCN further combines temporal convolution operation, which can simultaneously consider the spatial neighbor information and temporal dependency of air formation targets. Compared to GCN, STGCN is able to better extract and fuse the behavioral patterns of individual single targets within the formation over time, which helps to improve the accuracy of intention recognition, thus outperforming GCN to 87%. GAT introduces an attention mechanism based on GCN, which allows the model to adaptively assign weights to different targets and their neighbors within the air formation, focusing more on the key targets and their synergistic relationships within the formation and ignoring irrelevant or noisy interferences, which improves the recognition accuracy to 89%. The proposed STIRNet model combines the spatio-temporal attention mechanism and the spatio-temporal convolution operation to achieve the dual optimization in both time and space domains. Through spatio-temporal attention, the model not only handles the spatial interactions of formations, but also flexibly captures and enhances spatio-temporal features that are crucial for recognizing combat intention. As a result, STIRNet performs best in dealing with the complex spatio-temporal interaction behaviors of air formation targets, achieving the highest accuracy of 92%.

TABLE 3. Robustness evaluation results.

Standard Deviation	Accuracy (%)	Macro-Precision (%)	Macro-Recall(%)	Macro-F1(%)
	92	93	93	93
0.1	92	93	92	92
0.2	91	90	90	91
0.3	88	88	87	89

TABLE 4. Air formation target operational intent recognition results.

	Precision(%)	Recall(%)	F1(%)
Attack	98	100	99
Bombing	90	73	81
Refueling	96	98	97
Patrol	100	98	99
Transport	77	89	83
Scout	94	98	96
accuracy			92
macro avg	93	93	92

D. ROBUSTNESS ASSESSMENT

In order to evaluate the robustness of STIRNet, we add Gaussian noise with standard deviation of 0.1, 0.2 and 0.3 respectively to the key features such as coordinate value, velocity and altitude in the target data, and the results are shown in Table 3. From the table, it can be seen that when adding Gaussian noise with a standard deviation of 0.1, the four evaluation indexes of Accuracy, Macro-Precision, Macro-Recall and Macro-F1 are reduced by 0.54% on average; when adding Gaussian noise with a standard deviation of 0.2, the four evaluation indexes are reduced by 1.66% on average; when Gaussian noise with a standard deviation of 0.3 is added, the four evaluation metrics are reduced by 4.55% on average. This shows that when adding Gaussian noise with a standard deviation of 0.1, the performance of the model for formation target combat intention recognition is almost unchanged; when adding Gaussian noise with a standard deviation of 0.2, the recognition performance of the model decreases slightly; when adding Gaussian noise with a standard deviation of 0.3, the recognition performance of the model starts to show a more obvious decrease. The robustness evaluation reveals that STIRNet has strong robustness when processing target data. This is due to the fact that on the one hand STIRNet captures spatial dependencies through spatial convolution, and at the same time combines with temporal convolution to capture the dynamic characteristics of target data, which helps to understand and learn the complex interactions between air formation targets and the change rule over time, and even if it is interfered with by noise to a certain extent, it can still recognize the combat intention; on the other hand, the model can be used through the spatio-temporal attention mechanism to focus on features that carry critical information and are less affected by noise, thus improving the stability of the overall performance.

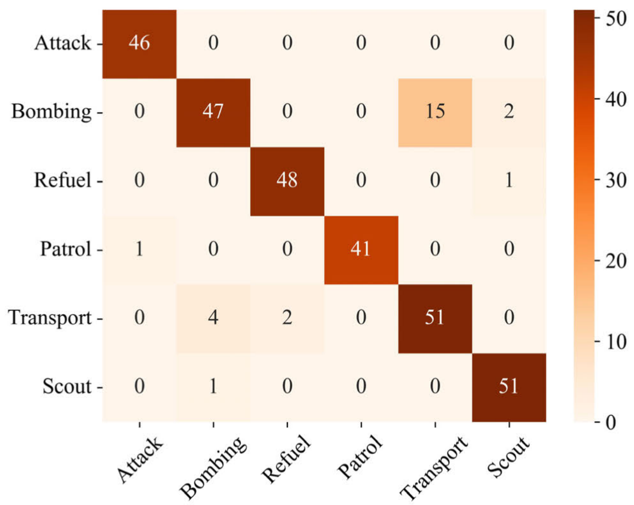


FIGURE 5. Confusion matrix for combat intention recognition of air formation targets.

E. THE RESULTS OF INTENTION RECOGNITION

The proposed STIRNet model is used to recognize the combat intention of air formation targets, and the confusion matrix of the recognition results is shown in Figure 5. From the figure, it can be seen that it is easy to be misjudged as transport when recognizing the combat intention of bombing, which is mainly due to the fact that the state and formation composition of the two are relatively close to each other when the air formation target is performing the relevant combat mission. When the combat intention of the air formation target is attack, the model is able to recognize all of them correctly. The results of air formation target combat intention recognition based on the STIRNet model are shown in Table 3, from which it can be seen that the model's precision is highest when the air formation target's combat intention is patrol; the model's recall is highest when the air formation target's combat intention is attack; the model's F1 is highest when the air formation target's combat intentions are attack and patrol; and the model's overall accuracy rate is 92%.

VI. CONCLUSION

We solve the complex spatio-temporal pattern recognition problem of air formation target combat intention recognition by the proposed STIRNet model. In this study, we first analyze the activity state of air formations and determine the main recognition framework; then we construct a spatio-temporal target intention recognition network STIRNet for air formations, in which, through the spatio-temporal attention mechanism, the model is able to adaptively learn and assign weights to key time steps and formation members, so as to focus on the momentary information that is crucial for recognizing the combat intention and the key target in the formation information, and through spatio-temporal convolutional operations, the model can capture these spatio-temporal correlations to effectively understand the behavioral patterns and their evolutions among targets within the formation. Comparing and analyzing with existing target intention recognition models, the STIRNet model we proposed demonstrates the strongest performance in understanding and dealing with complex spatial and temporal dynamic behaviors, and obtains the highest combat intention recognition accuracy. Through ablation experiments, the impact of each component in the STIRNet model on the overall intention recognition performance is analyzed and evaluated, and the results demonstrate that the model is able to more adequately mine and utilize the key spatial and temporal information in the data. We propose an innovative solution for the specific field of combat intention recognition of air formation targets, which provides technical support for command and decision-making.

To further expand the applicable fields of the model, considering the spatio-temporal data analysis capability of STIRNet, on the one hand, the model can be applied to civil airspace management, which can predict the flight paths and potential conflict points of the aircrafts through real-time analysis of a large amount of flight data, thus helping to optimize the route planning, improve the utilization rate of the airspace, and ensure the safety of flights. On the other hand, STIRNet can be used for UAV (Unmanned Aerial Vehicle) swarming behaviors research. By abstracting and processing the spatial and temporal interaction information between UAVs, it can accurately predict the behavioral patterns and development trends of UAV formations, which is of great significance for realizing efficient, orderly and safe UAV swarming operations. The research and practice of applying STIRNet to the above scenarios not only expands the performance boundaries of the model, but also foretells its broad prospects in solving complex dynamic system problems in other fields. In our future work, we will further explore the adaptive modification of STIRNet in these domains and carry out relevant experimental validation.

REFERENCES

[1] Z. Yang, P. Shi, T. Zhou, and W. L. Li, "Intention recognition method of space non-cooperative target based on fuzzy reasoning," *J. Beijing Univ. Aeronaut. Astronautics.*, pp. 1–18, 2024. [Online]. Available: https://kns.cnki.net/kcms/detail/11.2625.V.20240111.1643.002.html

[2] P. Ding and Y. F. Song, "An approach to air target intention recognition based on FCN-BiGRU," *Aero Weaponry*, vol. 30, no. 5, pp. 57–65, 2023.

[3] J. Xia, M. Chen, and W. Fang, "Air combat intention recognition with incomplete information based on decision tree and GRU network," *Entropy*, vol. 25, no. 4, p. 671, Apr. 2023, doi: 10.3390/e25040671.

[4] C. H. Deng, "Research on collective motion analysis and recognition," Ph.D thesis, Huazhong Univ. Sci. Technol., Huazhong University of Science and Technology, Wuhan, China, 2016.

[5] R. Yang, H. Gao, X. Wu, and Y. Feng, "An intention recognition method based on Bayesian network for UAV formation in confrontation," in *Proc. Int. Conf. Auton. Unmanned Syst.*, 2022, pp. 2799–2809, doi: 10.1007/978-981-99-0479-2_259.

[6] R. Yang, J. L. Yang, X. F. Liu, Y. L. Zhang, and Y. Y. Yan, "Air-ground cooperative operations intention recognition based on dynamic series Bayesian network," *Command Control Simul.*, pp. 1–11, 2024. [Online]. Available: https://kns.cnki.net/kcms/detail/32.1759.TJ.20240109.1104.008.html

[7] J. C. Yue, "Research and implementation of sea group target behavior intention reasoning technology based on big data," M.S. thesis, North China Inst. Comput. Technol., China Academic Electron. Inf. Technol., Beijing, China, 2022.

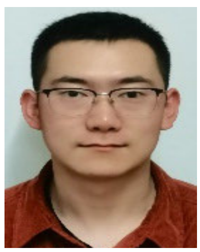
[8] Y. H. Wang, B. L. Dong, C. Gong, Z. Z. Shang, and K. N. Yao, "Dynamic threat assessment of air swarm targets based on intent recognition," *Comput. Modernization*, vol. 39, no. 12, pp. 100–104 and 111, 2023.
[9] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.
[10] V. S. Lalapura, J. Amudha, and H. S. Satheesh, "Recurrent neural networks for edge intelligence: A survey," *ACM Comput. Surveys*, vol. 54, no. 4, pp. 1–38, May 2022.
[11] D. Zhu and D. Wang, "Transformers and their application to medical image processing: A review," *J. Radiat. Res. Appl. Sci.*, vol. 16, no. 4, Dec. 2023, Art. no. 100680.
[12] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, and H. Ghayvat, "CNN variants for computer vision: History, architecture, application, challenges and future scope," *Electronics*, vol. 10, no. 20, p. 2470, Oct. 2021.
[13] D. Hupkes, M. Giulianelli, V. Dankers, M. Artetxe, Y. Elazar, T. Pimentel, C. Christodoulopoulos, K. Lasri, N. Saphra, A. Sinclair, D. Ulmer, F. Schottmann, K. Batsuren, K. Sun, K. Sinha, L. Khalatbari, M. Ryskina, R. Frieske, R. Cotterell, and Z. Jin, "A taxonomy and review of generalization research in NLP," *Nature Mach. Intell.*, vol. 5, no. 10, pp. 1161–1174, Oct. 2023.
[14] A. Sperduti and A. Starita, "Supervised neural networks for the classification of structures," *IEEE Trans. Neural Netw.*, vol. 8, no. 3, pp. 714–735, May 1997.
[15] M. Gori, G. Monfardini, and F. Scarselli, "A new model for learning in graph domains," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2005, pp. 729–734.
[16] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and deep locally connected networks on graphs," 2014, *arXiv:13126203*.
[17] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.
[18] R. Levie, F. Monti, X. Bresson, and M. M. Bronstein, "CayleyNets: Graph convolutional neural networks with complex rational spectral filters," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 97–109, Jan. 2019.
[19] R. Li, S. Wang, F. Zhu, and J. Huang, "Adaptive graph convolutional neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.
[20] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2014–2023.
[21] J. Atwood and D. Towsley, "Diffusion-convolutional neural networks," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 2001–2009.
[22] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.
[23] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–12.
[24] H. Dai, Z. Kozareva, B. Dai, A. Smola, and L. Song, "Learning steady-states of iterative algorithms over graphs," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1106–1114.
[25] Y. Li, R. Zemel, M. Brockschmidt, and D. Tarlow, "Gated graph sequence neural networks," in *Proc. ICLR*, 2016, pp. 1–19.
[26] D. D. Johnson, "Learning graphical state transitions," in *Proc. Int. Conf. Learn. Represent.*, 2016, pp. 1–19.
[27] N. Li and Z. Wang, "Spatial attention guided residual attention network for hyperspectral image classification," *IEEE Access*, vol. 10, pp. 9830–9847, 2022, doi: 10.1109/ACCESS.2022.3144393.
[28] D. Cheng, X. Gu, C. Qian, C. Du, and J. Wang, "Vehicle trajectory prediction with interaction regions and Spatial–Temporal attention," *IEEE Access*, vol. 11, pp. 130850–130859, 2023, doi: 10.1109/ACCESS.2023.3335091.
[29] H. T. Phan, N. T. Nguyen, and D. Hwang, "Aspect-level sentiment analysis: A survey of graph convolutional network methods," *Inf. Fusion*, vol. 91, pp. 149–172, Mar. 2023, doi: 10.1016/j.inffus.2022.10.004.
[30] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:160902907*.
[31] W. Zhou, J. Zhang, N. Gu, and G. Yan, "Recognition of combat intention with insufficient expert knowledge," in *Proc. 3rd Int. Conf. Comput. Model., Simul. Appl. Math., Wuhan, China*, 2018, pp. 27–28.
[32] W. W. Zhou, P. Y. Yao, J. Y. Zhang, X. Wang, and S. Wei, "Combat intention recognition for aerial targets based on deep neural network," *Acta Aeronautica et Astronautica Sinica*, vol. 39, no. 11, pp. 195–203, 2018.

CHENHAO ZHANG was born in Xiangyang, Hubei, China, in 1996. He received the M.S. degree in information and communication engineering from the Air Force Early Warning Academy, Wuhan, China, in 2020, where he is currently pursuing the Ph.D. degree. His research interests include situational assessment, data mining, and artificial intelligence.

YAN ZHOU received the Ph.D. degree in control science and engineering from Huazhong University of Science and Technology, Wuhan, China, in 2000. He is currently a Professor with the Air Force Early Warning Academy, Wuhan. His research interests include pattern recognition, data fusion, and image processing.

HONGQUAN LI received the Ph.D. degree in control science and engineering. He is currently a Professor with the Air Force Early Warning Academy, Wuhan. His research interests include data analysis, big data, and information systems.

YING XU received the bachelor's degree in automation engineering and the M.S. degree in information and communication engineering from the Air Force Early Warning Academy, Wuhan, China. She is currently an Associate Professor with the Air Force Early Warning Academy. Her research interests include data analysis, target detection, image recognition, and artificial intelligence.



YISHUAI QIN received the bachelor's degree in electronic information and engineering from Wuhan University of Science and Technology, Wuhan, China, in 2018, and the master's degree in information and communication engineering from the Air Force Early Warning Academy, Wuhan, in 2020. His research interests include signal processing, system engineering, and optimization algorithm.

LIANG LEI received the bachelor's degree in microelectronics from Lanzhou University, Lanzhou, China, in 2009. He is currently pursuing the master's degree in information and communication engineering with the Air Force Early Warning Academy. His research interests include data analysis, target recognition, and big data.