IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 November 2023, accepted 4 December 2023, date of publication 8 December 2023,
date of current version 27 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3341154

RESEARCH ARTICLE

Robust Air Target Intention Recognition Based on Weight Self-Learning Parallel Time-Channel Transformer Encoder

ZIHAO SONG, YAN ZHOU, WEI CHENG, FUTAI LIANG, AND CHENHAO ZHANG
Intelligence Department, Air Force Early Warning Academy, Wuhan 430019, China

Corresponding author: Yan Zhou (yunshanlele@sina.com)

This work was supported in part by the Graduate Student Sponsored Projects under Grant JY2022A017.

ABSTRACT Most existing air target intention recognition methods use only single-moment information, risking failure when acquiring data containing noise and many outliers. The robustness of methods that utilize continuous moment information has yet to be explored. This paper designs a robust recognition method for air target intention to address the above problems. The method takes data with noise and outliers as the object, based on a parallel time-channel Transformer Encoder and a weight self-learning unit. First, a detailed introduction to air target intention recognition and robust recognition is given, and the intention space and feature space are defined. Subsequently, the data samples are reconstructed using a fixed-step sliding window to increase the information utilized with multi-moment information as input. Finally, step-wise and channel-wise correlations are extracted using a time-axis Transformer Encoder and a channel-axis Transformer Encoder, respectively, and the weights of the two branches' outputs are automatically learnt using a weight self-learning unit. This enhanced self-attention network allocates attention weights between elements in the time and channel domain sequences to capture their long-range and short-range relationships and extract recognizable representations, making it robust to outliers and noise. The experimental results show that the model's recognition accuracy and composite F1 score reach 96.9% and 0.9676, and its performance remains well when the noise level and outliers proportion increase. The ablation and comparison experiments show its advantage in accuracy over other models.

INDEX TERMS Air target, intention recognition, transformer encoder, weight self-learning unit, multi-head attention.

I. INTRODUCTION

In many fields, it is of great significance to identify the intention of agents and their affiliated devices, which can help enhance cooperation and coordination among members of our group to form greater synergy, and can detect potential threats posed by enemy agents. Intention recognition is a typical pattern recognition problem that is often confused with activity recognition, as both rely on time series to tackle classification problems about agents' behavior. It is necessary to note that activity recognition focuses more on identifying the behavior that the agent has completed, while intention recognition is more concerned with identifying the agent's intention during the process of the agent's action, so that we have enough time to respond, that is, identifying its intention before its purpose is achieved. The normalized definition of intention recognition was first given by Kautz and Allen [1]. After that, intention recognition has attracted widespread attention in human-computer interaction [2], [3], recommendation system [4], [5], [6], pedestrian trajectory prediction [7], [8], [9], and driver lane change prediction [10], [11], [12]. Existing intention recognition methods can be divided into model-based and data-based methods. The former predefines the model and dynamically adaptively adjusts the relevant parameters, and finally forms a deterministic model to carry out the discriminant intention. The latter is based on machine

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

learning classifiers and neural networks, learning hidden patterns directly from data or features to complete intention identification.

In the military and defense fields, the recognition and understanding of enemy intentions is also of great concern. As the core activity of battlefield situation cognition and understanding, identifying enemy target intentions can effectively enhance battlefield awareness and decision-making efficiency [13], [14]. And in information and intelligent warfare, air targets, as very active, highly threatening, with great uncertainty and combat capability potential, are important objects of intention recognition. The accurate recognition of air target intentions can enable our side to enhance understanding of the situation and gain the initiative in the air battlefield [15], [16]. With the development of military technology and aerospace weapons, the confrontation and complexity of air defense battlefields have increased significantly, showing the following characteristics: air target types and numbers have increased sharply, the amount of data obtained and processed by sensors has exploded, the intensity of confrontation and the degree of the game are significantly improved, the decision-making time has been greatly shortened, the electromagnetic environment is more complex, and the desirability of data is difficult to guarantee [17], [18]. In this case, in the face of a large amount of data with noise and outliers, it is difficult for back-office technicians and front-office commanders to quickly and correctly extract the hidden key situation elements from them, and then infer the intention. Therefore, there is an urgent need to use automated and intelligent means to achieve air target intention recognition.

In the past, researchers have mostly focused on the problem of target activity identification in air situation understanding [19], [20], [21], [22]. Undoubtedly, activity recognition is of great help in enriching intention recognition knowledge. However, real-time inference of intentions has stronger military application value. Recently, there has been a proliferation of work on intention recognition as the need for and awareness of in-event intention recognition has increased. By analyzing a large number of works on air target intention recognition, we found that: (1) A great deal of them only utilize state information at a single moment in time, and information prior to that moment is not applied. Generally, the intention of the target often needs to be reflected through continuous stable behavior or behavior changes over a while. Hence, it is one-sided to use only the information at the current moment for intention recognition [23], [24], [25]; (2) Methods that utilize multi-moment information depend on the perfect data assumption and do not investigate performance when there are imperfections in the data, such as outliers and noise. In fact, within the intricate adversarial landscape of modern warfare, the presumption of flawless data is untenable.

In this study, our focus is on the issue of intention recognition in the presence of noisy and outlier-laden data. Specifically, we propose a robust intention recognition model based on multi-moment information features and an improved self-attention-based network. The model is named as WSPTCTE-IR. WSPTCTE refers to weight self-learning parallel time-channel Transformer Encoder, IR refers to intention recognition. Our contributions are as follows:

(1) We describe and analyze the problem of robust air target intention recognition. Based on the general framework for intention recognition, this paper summarizes solution paths for robust recognition. Additionally, due to the instability and uncertainty of non-numerical target features, the feature space is restructured to be dominated by numerical target features.

(2) We develop a robust data-driven end-to-end model for recognizing intention amidst data containing noise and outliers based on WSPTCTE. The model is designed to operate effectively without the need for outlier or noise processing. Here, the fixed-step sliding window is utilized to reconstruct the intention recognition features in the time axis to increase the data used. Additionally, we utilize the self-attention mechanism on the channel and time dimensions to acquire more extensive global and local correlations. This approach facilitates the extraction of implicit information in both temporal and channel domains. And the introduced weight self-learning unit can adaptively learn the weights of the two parallel branches' outputs to avoid performance decay when brutely concatenating them. To the best of our knowledge, this is the first time that a time-channel self-attention-based network has been applied to air target intention recognition.

(3) We conduct a large number of comparative experiments, robustness tests, and ablation experiments to explore the influence of sliding window length, batch size, learning rate, epoch, dropout probability value, noise level, the proportion of outliers, and other factors on the intention recognition results, which can prove the model's effectiveness and robustness and provide very beneficial references for subsequent research.

II. RELATED WORKS

Air target intention recognition is a cognitive activity that analyses and identifies the air target's combat intention in the game process, and finally forms an identification conclusion for further analysis by technicians and then assists commanders in decision-making. It is the focus of the air battlefield situation understanding and analysis and a hotspot in current research. This section introduces model-based and data-based intention recognition methods, and the pros and cons of both are explained.

A. MODEL-BASED METHODS

The model-based method mainly includes template matching, expert system, decision theory, Bayes network, etc. Xia analyzed the situation knowledge base and event association in intention recognition reasoning, further studied the matching inference framework of intention recognition, proposed an intention recognition template matching method based on Dempster-Shafer evidence theory, and illustrated the possibility of the method to identify and judge the target intention with an example [26]. Aiming at target tactical intention recognition in ship command decision-making and according to the characteristics of domain knowledge, Leng et al. proposed an algorithm for the support degree of target real-time state to the intention type based on the similarity of feature components, and the evidence theory is used to integrate the support degree of each moment to form a sequential recognition approach of target tactical intention [27]. Li et al. constructed a template-based intention recognition reasoning framework, studied the situation estimation inference algorithm, and proposed a general template matching algorithm for situation estimation [28]. Template matching methods are simple to implement and conform to basic human cognition. However, establishing its template database relies heavily on expert knowledge, and it is difficult to be competent in intention reasoning in the case of undesirable data.

Expert systems are intelligent computer program systems that contain a large number of knowledge and experience of experts in a specific field. It can apply artificial intelligence and computer technology to imitate human experts' decision-making process based on the system's knowledge and experience to carry out reasoning and judgment and further solve complex problems that require human experts to cope. Moreover, domain knowledge bases and inference frameworks or models are at their core. Song et al. constructed a reasoning decision support system for target intention characterized by an expert system and established an intention hierarchical reasoning framework by utilizing decision trees and a data-driven reasoning control mechanism based on the distributed characteristics of intention reasoning input information and the hierarchical decomposition of intention [29]. Wu and Li proposed a model for determining air target attack intention based on intuitionistic fuzzy generative rule reasoning and multi-attribute decision-making theory, which avoids the problem of combinatorial explosion of expert knowledge due to excessive battlefield information and ensures the computational speed of the system [30]. Expert systems are capable of knowledge representation and computational reasoning. However, abstracting a complete knowledge base and inference rules makes them more challenging to implement, less fault-tolerant, and less capable of learning. In an information and intelligent battlefield with a complex electromagnetic environment, it is difficult to rely on mechanical rules of reasoning to summarize the complex evolution of the situation and achieve an accurate understanding of target intentions.

There has also been much work using decision theory to achieve intention recognition of air targets. Most of this work model intention recognition as a multi-attribute decision problem. Li proposed a sequential three-branch decision-making method based on the characteristics of delayed decision-making and the time-series relevance of air target combat intention recognition. By establishing a mathematical model combining multi-category three-branch decision-making, sequential ideas, and target intention recognition, the intention recognition process is divided into several stages on the timeline, and the three-branch decision-making-based air target intention recognition model is used to obtain the intention recognition results of the target in the current stage [31]. Yang proposed a cost-sensitive multi-category three-branch decision-based method for air target intention recognition. The method calculated the intention with the lowest misclassification cost loss value at each recognition stage and obtained a remarkable recognition result, thus avoiding conflicting results. Simultaneously, the method solves the non-recognition problem caused by the missing delay domain of multi-category three-branch decision methods [32]. Intention recognition based on multi-attribute decision methods has a solid mathematical basis. However, it is cumbersome and has a high risk of failure when the decision variables are undesirable.

The Bayesian network, derived from the Bayes Rule, has been widely used to solve uncertainty problems. The steps involved in using the Bayesian network to recognize target intentions can be briefly summarized as follows: constructing the network, determining the parameters, updating the parameters, and outputting the results at the final [33]. Yue proposed a dynamic Bayesian network model-based behavioral intention inference method based on time, space, target, event, and mission knowledge elements. By analyzing the decomposition and execution process of behavioral intention, a sequential Bayesian network model was established to describe the behavioral intention planning and analyzing process, which can complete the intention inference of group targets in a complex naval battlefield environment [34]. Qing et al. proposed an optimized Bayesian network algorithm for air swarm target combat intention recognition by extracting external features of target swarm data chains as network nodes, and the effectiveness of the algorithm was verified through simulation [35]. Xu et al. introduced information entropy to optimize dynamic sequential Bayesian networks to objectively assign attribute weights by analyzing the amount of helpful information from different participating attributes to identify air target combat intention effectively [36]. The Bayesian network has strong causal probabilistic inference capabilities, allowing inferences from incomplete, imprecise, or anomalous information segments. However, it has great difficulty in determining the prior and conditional probabilities at each node, which limits their application to some extent.

B. DATA-BASED METHODS

Unlike model-based approaches, data-based approaches are data-centric and rely on machine learning and deep learning techniques to improve intention recognition performance. Meng used support vector machine (SVM) and 19 low correlation features to identify the most concerned multi-aircraft coordinated air warfare attack intention. The method also combined the use of dynamic Bayesian network, radar models, and threat assessment models to extract key features that can be utilized for generic intention recognition, resulting in a considerable improvement in the accuracy [37]. Hu et al. built an air target intention recognition model based on the random forest (RF) algorithm, and the results showed that its recognition accuracy has advantages over other algorithms [38]. Yang et al. proposed a cascaded SVM-based online phased recognition method for the tactical intention of over-the-horizon air combat targets, constructed a progressive identification model from target maneuver elements, tactical maneuver behavior to the tactical intention [39]. Wang and Li proposed an XGboost-based target intention recognition method to improve the accuracy, which ultimately relies on D-S evidence theory to output sequential intention probabilities [40]. Machine learning methods are theoretically well-grounded and excel at intention recognition with complete data. However, their inability to extract deep features from large-scale data has limited potential for further application when facing extensive undesirable data.

Over the past decade, deep learning algorithms have been applied to computer vision, natural language processing, recommendation systems, time series analysis, and other fields, achieving remarkable results [33]. By using deep neural networks (DNNs) to learn and process intention features layer-by-layer, the high-level information of the battlefield situation can be progressively extracted from the shallow features. Xue et al. proposed a method for intention recognition of air targets based on convolutional long and short-term memory (LSTM) networks, which combines the temporal feature extraction capability of LSTM layers with the local feature mining capability of convolutional neural networks (CNNs) to improve recognition performance [41]. Teng et al. constructed a deep neural network for air target intention recognition, which improves the accuracy of recognition by using an attention mechanism to assign weights to each attribute prior to the backbone network [42]. The above approaches provide instrumental explorations using deep learning methods. However, they treat intention recognition as a post-event analysis activity, which tends to confuse it with activity recognition. Intention recognition should be more of an in-event analysis activity. Some scholars viewed it as an in-event analysis activity, and several recognition algorithms using deep learning methods were proposed. Qu fed the critical motion state information and the corresponding labels into the designed fully connected network (FCN), CNN, and LSTM. The experimental results showed that the LSTM-based recognition model achieved the best results [43]. Wang et al. proposed a hybrid neural network-based quick-in-event intention recognition model using neural network modules adapted to different data types [44]. Focusing on the requirements of timeliness and interpretability of air target intent recognition, Wang proposes a method based on the bidirectional gate recurrent unit (BiGRU) and conditional random field. It can provide more accurate recognition results at any time [33]. Wang et al. proposed a real-time target tactical intention recognition algorithm based on bi-directional long short-term memory (BiLSTM); the simulation results show the effectiveness [45].

Of all the available deep-learning-based methods, the RNN is the most commonly used backbone network. RNN's recursive structure enables it to process temporal information in sequences. However, this structure can lead to gradient vanishing and long-term dependency issues, which means that researchers tend to use smaller input lengths when employing the RNN for intention recognition. Unfortunately, such an approach curtails the amount of information that can be utilized. Furthermore, it is challenging for RNNs to prioritize the correlation information among attributes, and their sequential operations result in computational inefficiency. Furthermore, the above work does not analyze the robustness of their proposed method in the presence of noise and outliers in the data obtained.

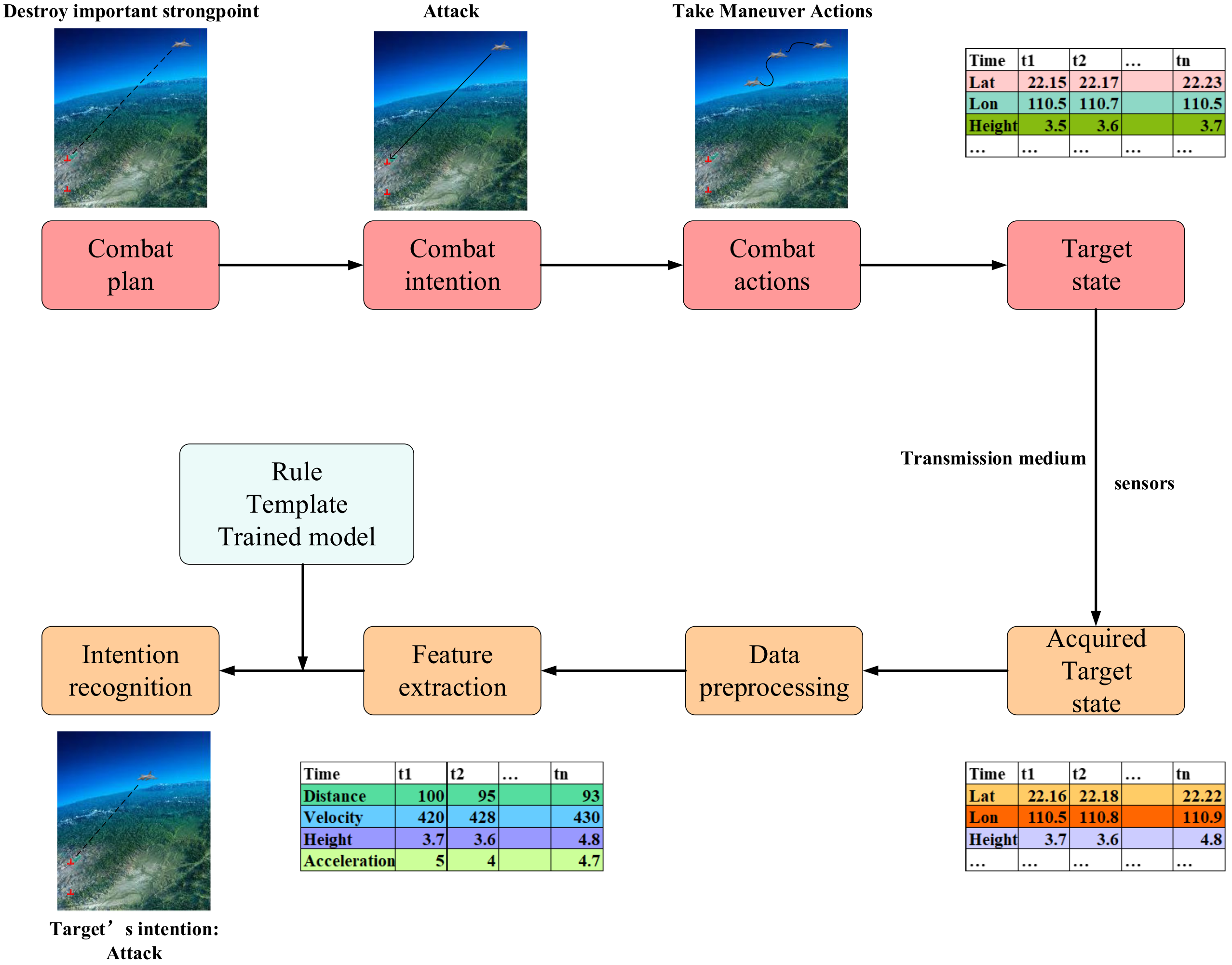
To deal with these limitations and effectively recognize the intention of air targets when noise and outliers exist in the data acquired, we choose the deep learning and construct a robust recognition model based on WSPTCTE. The enhanced self-attention network, WSPTCTE, allocates attention weights between elements in the time and channel domain sequences to capture their relationships and generate outputs, making it robust to outliers and noise. Additionally, the self-attention mechanism can more effectively capture both long and short-term dependencies and has a greater capacity for feature extraction. In the next section, the robust intention recognition of air targets was described in detail. After that, we introduce our proposed WSPTCTE-IR model.

III. ROBUST AIR TARGET INTENTION RECOGNITION

A. THE DESCRIPTION

Air target intention recognition is analyzing and identifying the intention of an air target based on military domain knowledge in a real-time, hostile battlefield environment using information about the air target's state gathered from sensors. It differs from activity recognition in that it focuses on continuous identification during the event so that we can react to the enemy target's intentions quickly [46], [47]. The intention often represents the enemy's operational plans and implicit mappings of the enemy's mindset that are not directly accessible and available through data and are difficult to describe. However, the enemy targets must have the appropriate location, speed, and other characteristics to fulfil the operational intent and thus advance the operational plan. In other words, the enemy must be guided by plans to achieve intentions through reasonable actions and states that can be detected, which is the most fundamental basis for our recognition of intentions.

In the information and intelligent battlefield, the electromagnetic environment is complex and volatile, and the data acquired by non-cooperative receivers often contains uncertainty that includes noise and outliers, in addition to data fluctuating within an acceptable range. Robust intention



Time	t1	t2	…	tn
Lat	22.15	22.17		22.23
Lon	110.5	110.7		110.5
Height	3.5	3.6		3.7
…	…	…	…	…

Time	t1	t2	…	tn
Distance	100	95		93
Velocity	420	428		430
Height	3.7	3.6		4.8
Acceleration	5	4		4.7

Time	t1	t2	…	tn
Lat	22.16	22.18		22.22
Lon	110.5	110.8		110.9
Height	3.7	3.6		4.8
…	…	…	…	…

FIGURE 1. **The data flow for air target intention generation and recognition. In this figure, In the figure, we take an example of an enemy target carrying out an attack intention.**

recognition is concerned with constructing and designing algorithms and models to achieve high-level recognition accuracy when the data captured by sensors contain noise and significant outliers.

Air targets have always attracted much attention as very active, highly threatening elements and have great uncertainty and combat capability potential. Moreover, with the development of military technology and aerospace weapons, the confrontation and complexity of the air defense battlefield have increased significantly, making the air target intention data we acquired contain outliers and noise, which makes some recognition methods decline in performance or even fail. Therefore, there is a very urgent need and a promising military application for robust recognition of air target intention. Here we first give the data flow for air target intention generation and recognition in Figure 1.

The enemy's operational plans lead to the creation of corresponding intentions, as can be seen in Figure 1. To achieve that intention, the enemy target needs to launch actions. Guided and driven by different operational intentions, the state information will inevitably diverge, and this divergence is the fundamental basis for intention differentiation. Steps such as data acquisition, pre-processing, feature extraction, and finally, recognition using rules, templates, or offline trained models are performed sequentially on the receiving side (in this case, on our side).

Formally, air target intention recognition can be described as mapping the target state information acquired by our sensors to the enemy's operational intention. In complex electromagnetic environments, the air target information collected may contain noise and significant outliers, rendering the data unusable at some point. They cannot be relied upon to infer intention.

The usual idea for this problem is to construct some specialized work in the data pre-processing and feature extraction phase:

(1) In data pre-processing: for noisy data, the denoising algorithm is used to improve the signal-to-noise ratio; for data with significant outliers, the anomaly detection algorithm is used to detects and smooths the outliers.

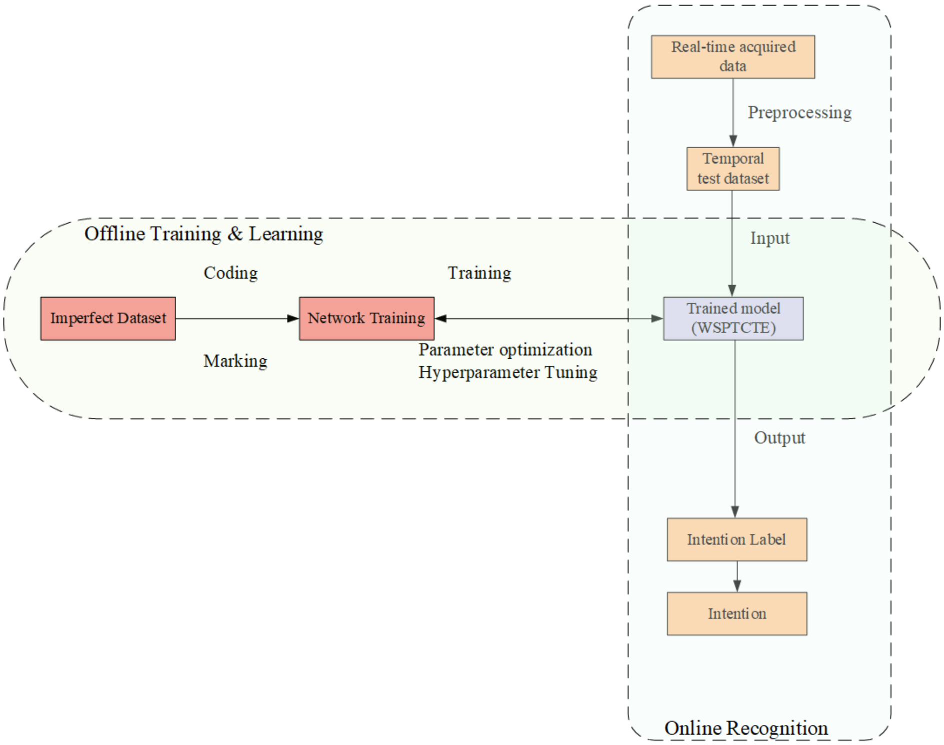


FIGURE 2. The intention recognition framework.

(2) In feature extraction: specialized work at the data pre-processing stage is not always effective when faced with data containing noise and significant outliers. Therefore, more robust features need to be prudently extracted to ensure performance.

Moreover, another solution is not to do special processing on data with noise and outliers, and to build a recognition model using target state information from several successive moments, i.e., to construct and learn the following implicit mapping relations:

Specifically, the set of features utilized at moment t is defined as $\boldsymbol{X}_t, \boldsymbol{X}_t = \boldsymbol{x}_{t-N+1}, \boldsymbol{x}_{t-N+2}, \cdots, \boldsymbol{x}_t\}, t \geqslant N$, where denotes the features at the $N-$ th moment before t; and the target's intention at moment t is denoted as, then the mapping of the target feature set to the target's intention can be denoted as $\boldsymbol{Y}_t = f(\boldsymbol{X}_t) = f(\{\boldsymbol{x}_{t-N+1}, \boldsymbol{x}_{t-N+2}, \cdots, \boldsymbol{x}_t\}), t \geqslant N$. And there is no doubt that such solutions require the design of algorithms or networks with high robustness and great learning capability.

Based on the above analysis, a robust air target intention recognition method based on multi-moment data information and WSPTCTE is proposed, which does not require anomaly detection, smoothing and de-noising operations on data with noise and outliers in the preprocessing stage. The framework is shown in Figure 2.

The method is divided into two stages: offline training & learning and online recognition. Offline training & learning refer to relying on a pre-organized air target intention dataset, following the general paradigm of deep learning methods, using the train set to learn and optimize the weight parameters, to finally obtain a trained intention recognition model which implicitly establishes a mapping relationship from the feature set to intention space. In the online recognition phase, the real-time data is pre-processed by normalization and coding and form the temporal test set; then the processed data is fed into the trained model in sequence to obtain real-time intention recognition results.

B. AIR TARGET INTENTION SPACE AND FEATURE SPACE

1) AIR TARGET INTENTION SPACE

The nature and granularity of intention vary with different operational contexts, weapon and equipment employment, and combat intensity. Therefore, an essential basis for identifying enemy intention is a reasonably prudent definition of the target's intention space based on the relevant operational context, the primary attributes and capabilities of the enemy's and our combat units, and the operational plan.

In this paper, we focus primarily on the air defense early warning operation. In this context, the target intention space is established by considering the attributes and tasks of the enemy targets. This intention space contains six intentions: {attack, anti-submarine, aerial refueling, police patrol, retreat, and airborne warning and control (AWAC)}, and a detailed description is shown in Table 1.

It is essential to clarify that in a dynamically changing air battlefield, there may be more than one intention of the target at a given time, and the intention of the target may also change

TABLE 1. The description of air target intentions.

Intention	Description
Attack	The target's state of movement changes rapidly, and at the right moment, it will fire offensive munitions against important targets
Anti-submarine	Detecting and tracking to deter underwater targets
Aerial refueling	Refueling fighters in the air
Police patrol	Conducting alert patrols in response to potentially uncertain events
Retreat	Evacuating quickly after completing tasks or when safety is threatened
AWAC	Detecting targets in the air and providing intelligence support to other operational entities

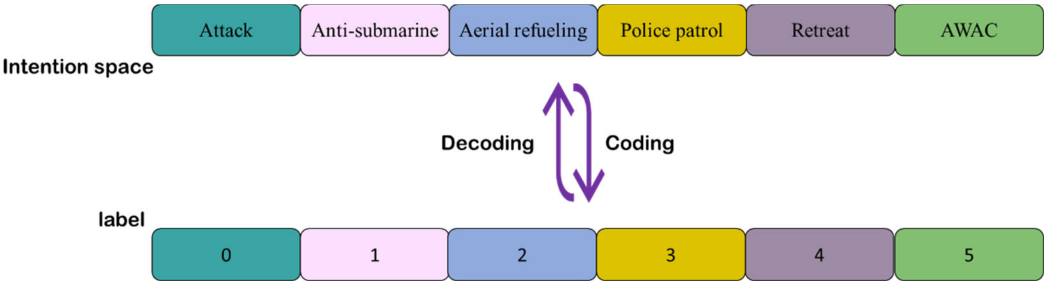


FIGURE 3. Intention coding and decoding.

in response to the battlefield situation and the missions of our fighters. The subsequent study in this paper assumes that each target has only one primary intention at a given time and uses it to label and identify the sample data.

Intention recognition is a multi-classification problem and therefore requires supervised learning during training. Moreover, the intention space of airborne targets is cognitive knowledge learned and stored by the human brain. In the offline training & learning phase, we need to translate that knowledge into digital labels of intention that the neural network can process and recognize. Label construction relies heavily on humans, i.e., the commander or operator extracts valuable features from battlefield situational data, follows rules of thumb, and identifies the enemy target's intention concerning the target's past activity patterns. The rule shown in Figure 3 is utilized to label the above six intentions, which facilitates the training and recognition of the model.

2) AIR TARGET INTENTION FEATURE SPACE

The air target's intention is reflected by its actions, status and the variation of both, which may vary considerably in different operational contexts and for the same operational intention. Therefore, we first give the operational scenario considered in this paper: enemy air targets take tactical actions against our entity, the target type, radar status, and electronic reconnaissance equipment status of the enemy are not available, and we can only rely on the trajectory of the fusion center to track and monitor the target, but the trajectory of the fusion center contains noise and outliers.

Since air targets will inevitably take specific tactical actions or maneuvers to achieve intentions when driven by a mission, the motion states of air targets in the time domain with different intentions and their evolution can be used as a basis for intention recognition. For example, aircraft with anti-submarine intention often fly at low altitudes and can be observed moving from medium to high altitudes to low altitudes and hovering around priority targets, and corresponding changes in altitude and speed, as well as nearly cyclical reciprocation of heading angle and azimuth, can be observed; AWAS aircraft will have to fly in an arc or oval shape over a range of altitudes to detect a target, which results in significant nearly periodic trends in heading angle, distance, and azimuth; aircraft with attack intention often use low altitude penetration, so the frequency of change in speed, height, acceleration, heading angle, azimuth and distance is high; aircraft with retreat intention will gradually get further away from to our entity.

Based on the above analysis and considering data availability, six motion state features, such as velocity, acceleration, height, heading angle, azimuth and distance, are chosen to build the intention feature set. Furthermore, the description of them is given in Table 2.

Here, $(x_{\text{target}}, y_{\text{target}}, z_{\text{target}})$ and $(x_{\text{op}}, y_{\text{op}}, z_{\text{op}})$ are the coordinates of the air target and our unit in the Cartesian coordinate system. It should be noted that the latitude, longitude, and height obtained from the fusion center need to be converted to Cartesian coordinate system coordinate values using the coordinate system conversion formula to obtain the above values more efficiently [25].

IV. RECOGNITION MODEL

In this paper, a robust recognition model for air target intention based on WSPTCTE and data containing noise and outliers is constructed. The steps of WSPTCTE-IR are shown below:

TABLE 2. The description of air target intention features.

Characteristic	Description	Formula or illustration
Distance (m)	The distance of the projection point on the ground of the air target from our entity	$d = \sqrt{\left(x_{\text{target}} - x_{\text{op}}\right)^2 + \left(y_{\text{target}} - y_{\text{op}}\right)^2}$
Velocity (m/s)	The velocity of the air target	$v = \frac{\Delta\sqrt{\left(x_{\text{target}} - x_{\text{op}}\right)^2 + \left(y_{\text{target}} - y_{\text{op}}\right)^2 + \left(z_{\text{target}} - z_{\text{op}}\right)^2}}{\Delta t}$
Acceleration ($\mathrm{m/s^2}$)	The acceleration of the air target	$a = \frac{\Delta v}{\Delta t}$
Heading angle (°)	The angle between the direction of the air target and the Earth's North Pole	θ_1, see in Figure 4
Azimuth (°)	The angle from our entity to the direction of the air target direction.	θ_2, see in Figure 4
Height difference (m)	The difference between the altitude of the air target and our entity's altitude	$\triangle H = \lvert z_{\text{target}} - z_{\text{op}} \rvert$

Step 1 Preprocess basic information and construct the dataset. Feature extraction and unified coding of air target trajectory information are performed to build a normative data set.

Step 2 The train and test sets are divided according to a specific ratio. The train set is normalized, followed by the normalization of the test set using the same parameters. Finally, the data is sliced with the same size sliding window and step to obtain the final train and test set. The train set is used for offline training & learning, and the test set is used for online recognition.

Step 3 A parallel time-channel Transformer Encoder network is built, using the self-attention mechanism to mine and extract step-wise and channel-wise correlations in time and channel dimension, respectively, to try to eliminate the harmful effects of noise and outliers; and a weight self-learning unit is conducted to automatically learn the weights of outputs of two branches to avoid possible performance degradation caused by direct concatenating.

A. DATA PREPROCESSING AND DATASET ESTABLISHMENT

First, we generated trajectories for six intentions in the combat simulation system based on principles of air warfare, the attribute and capability of air targets, and guidance from domain experts. We then extracted the motion state information from the raw data.

Noise and a variable percentage of outliers are added to the motion state data to bring the simulation data closer to the actual data acquired by the non-cooperative receiver during operations in complex electromagnetic environments, and the process can be expressed as:

$$M\left(t\right) = \left[d, v, a, \theta_1, \theta_2, \Delta H\right]^T + v\left(t\right) + u\left(t\right) \tag{1}$$

where $M\left(t\right)$ represents the feature vector at time t, $v\left(t\right)$ represents the Gaussian noise:

$$v\left(t\right) \sim N([0]_{6\times 1}, Q) \tag{2}$$

$$Q = \mathrm{diag}(\sigma_d^2, \sigma_v^2, \sigma_a^2, \sigma_h^2 \sigma_d^2, \theta_1^2, \theta_2^2) \tag{3}$$

σ represents the variances of corresponding Gaussian noise, $u\left(t\right)$ represents possible outliers at time t:

$$u\left(t\right) = [u_1, u_2, \ldots, u_6]^{\mathrm{T}}, \quad u_i \in \mathrm{R} \tag{4}$$

The dataset is then divided into a train set and a test set according to a specified scale, after which the data are normalized to remove the effects of unit and scale differences between the features: the train set data is first normalized using the min-max normalization method, and then the test set data are normalized using the same procedure and parameters:

$$x_{i,j,k}^{\text{Norm}} = \frac{x_{i,j,k} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \tag{5}$$

where $x_{i,j,k}$ represents raw values for the k-th sample under the j-th class of features, at the i-th sampling point, $x_j^{\min}$ and $x_j^{\max}$ denotes the minimum and maximum values of the j-th dimension of features in all samples in train set, $x_{i,j,k}^{\text{Norm}}$ represents normalized data.

In the online recognition phase, we must input feature segments with a fixed length in sequence to the trained model to obtain real-time intention recognition results. Here, a sliding window with a fixed step and length is used to slice the entire feature time series to assemble the data for subsequent model parameter learning in the offline training and online recognition. The feature input is a matrix $\boldsymbol{X}_t$ of $6\times s$,

$$\boldsymbol{X}\left(t\right) = [\boldsymbol{M}(t-s+1), \boldsymbol{M}(t-s+2), \ldots, \boldsymbol{M}(t)], s < t \leqslant T \tag{6}$$

where s represents the length of the sliding window, and T denotes the entire time series length. $\boldsymbol{M}\left(t\right)$ represents the feature vector of 6×1 at moment t.

B. THE WSPTCTE NETWORK

The general framework of the recognition network constructed in this paper is shown in Figure 5.

As can be seen from the figure, the input to the model is an $N\times s$ matrix with values in the range [0,1]. The network consists of two parallel branches, which are similar in structure but have distinctly different functions: the upper branch is time-axis Transformer Encoder (TTE) used to extract step-wise correlations, and the lower is channel-axis Transformer Encoder (CTE) used to extract channel-wise correlations. Subsequently, a weight self-learning unit (WSU) is introduced to learn the weights of the two branches'

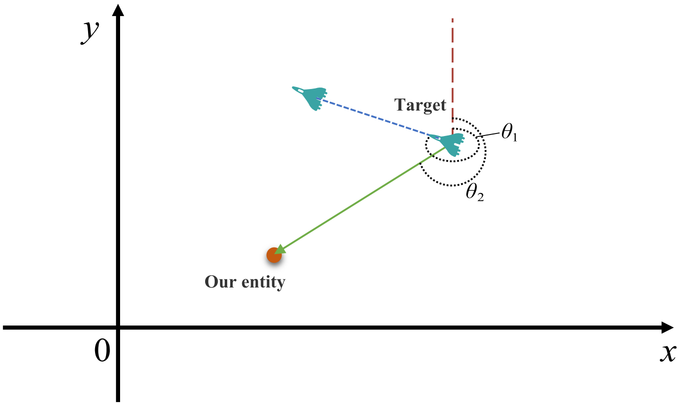


FIGURE 4. The illustration of the heading angle and azimuth.

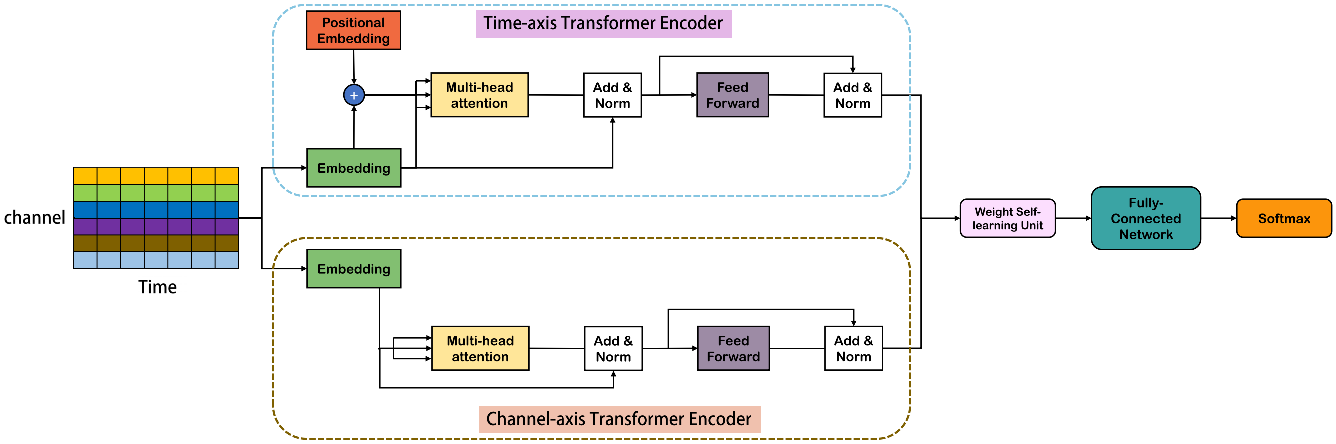


FIGURE 5. The general framework of the WSPTCTE.

outputs adaptively. Finally, the output elements of the two branches are multiplied with corresponding weights. The two are concatenated and fed into the fully connected layer (FCL) and the Softmax classifier to obtain the classification result.

This section introduces embedding, the Transformer Encoder, multi-head attention, the feed forward layer, and the weight self-learning unit.

1) EMBEDDING

The original Transformer uses learnable embeddings to transform input tokens into word vectors of dimension d_{model}; the purpose of the embedding layer is to reduce the dimensionality of the word vectors [48]. In this paper, the role of the embedding layer is to increase dimensionality to improve the discriminability of features in low-dimensional spaces. We simply change the embedding layer to the FCL, and the tanh function is added to replace linear projection,

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{7}$$

In addition, since the Transformer is hard to capture the natural sequential relation of the time step, it is necessary to fuse the positional encoding into the time-step features. Here a fixed positional encoding is utilized (only in the upper branch) and shown below:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{8}$$

$$PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{9}$$

where pos represents location information for each moment, d_{model} represents the dimension of the output of the embedding layer, $PE(pos, 2i) \in R^{d_{\text{model}}}$ and $PE(pos, 2i+1) \in R^{d_{\text{model}}}$ denotes the positional encoding when the dimension index is even and odd, $i = 0, 1, 2\cdots, \frac{d_{\text{model}}}{2} - 1$. In this paper, d_{model} is set to 512.

Moreover, it is essential to note that the position of the channels has no relative or absolute correlation with the input, as the input should have no change if we switch the order of the channels. Therefore, position encoding is only utilized in the upper branch.

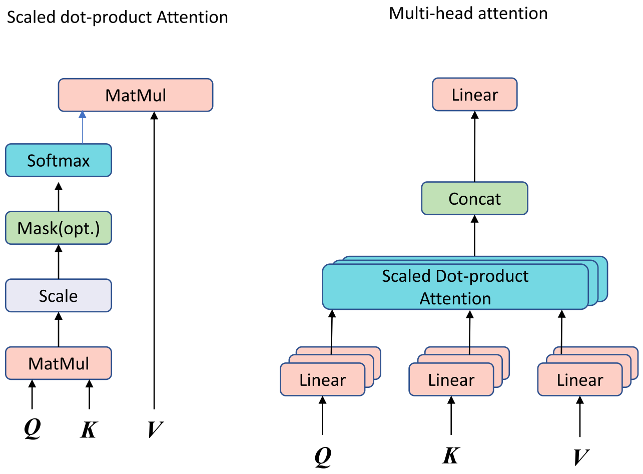


FIGURE 6. **Attention mechanism.**

2) THE TRANSFORMER ENCODER

In the WSPTCTE, the upper and lower Transformer Encoder both has two sub-layers. The first is a multi-head attention mechanism (MHA) layer, and the second is a feed forward network (FFN). And a residual connection is utilized around each sub-layer, followed by layer normalization (LayerNorm) operation [49], [50]. The output of each sub-layer is

$$\text{LayerNorm}(x + \text{sublayer}(x)) \tag{10}$$

where sublayer(x) refers to MHA or FFN. To facilitate these residual connections, all sub-layers and embedding layers in the model produce the same outputs of the same dimension d_{model}.

3) MULTI-HEAD ATTENTION

The Transformer Encoder uses MHA to capture long-range dependencies, and the heart of MHA is scaled dot-product attention that maps a query and key-value pair to an output where query, key, value, and output are all vectors, as shown in Figure 6.

$\boldsymbol{Q}$, $\boldsymbol{K}$and $\boldsymbol{V}$ represent query, key, value, and the dimensions of three are d_{model}, and the attention weight matrix is given by:

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_{\text{model}}}}\right)\boldsymbol{V} \tag{11}$$

$$\text{Softmax}\,(z_i) = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_c}} \tag{12}$$

where z_i represents the output of i-th node, C represents the number of output nodes.

Instead of performing a single attention function, MHA linearly project the queries, keys and values h times with different learned linear projections to d_k, d_k and d_v dimensions respectively. The process can be expressed as

$$\text{head}^x = \text{Attention}(\boldsymbol{Q}\boldsymbol{W}_q^x, \boldsymbol{K}\boldsymbol{W}_k^x, \boldsymbol{V}\boldsymbol{W}_v^x) \tag{13}$$

$$\text{MultiHead}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = (\mathop{\Big\Vert}_{x=1}^{h} \text{head}^x)\boldsymbol{W}_o \tag{14}$$

where $\boldsymbol{W}_{\text{q}}^x, \boldsymbol{W}_{\text{k}}^x \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $\boldsymbol{W}_{\text{v}}^x \in \mathbb{R}^{d_{\text{model}} \times d_v}$ and $\boldsymbol{W}_o \in \mathbb{R}^{h \cdot d_v \times d_{\text{model}}}$ denotes corresponding mapping parameter matrix, $\mathop{\Big\Vert}_{x=1}^{h} (\cdot)$ denotes the concatenating operation on heads, h denotes the number of attention heads. In this paper, we employ $h = 8$ heads, $d_k = d_v$ =64.

4) FEED FORWARD NETWORK

Feed forward network consists of two cascades of FCLs and a ReLu function, and the arithmetic process is as Equation (15)

$$\text{Feed Forward}\,(\boldsymbol{x}) = \boldsymbol{W}_2\max(0, \boldsymbol{W}_1 x + \boldsymbol{b}_1) + \boldsymbol{b}_2 \tag{15}$$

where $\boldsymbol{W}_1 \in \mathbb{R}^{d_{\text{model}} \times d_{hidden}}$ represents the first linear mapping parameter matrix, and $\boldsymbol{W}_2 \in \mathbb{R}^{d_{hidden} \times d_{model}}$ represents the second one. $\boldsymbol{b}_1 \in \mathbb{R}^{d_{hidden}}$, $\boldsymbol{b}_2 \in \mathbb{R}^{d_{\text{model}}}$ are bias vectors, and $\boldsymbol{x}$ represents the input. In this work, d_{hidden} is set to 1024.

5) WEIGHT SELF-LEARNING UNIT

One of the simplest ways to merge the upper and lower branch Transformer output features is to concatenate them directly, but this approach may degrade performance. Here a weight self-learning unit (WSU) is introduced to determine the weights of the upper and lower branches automatically:

1. The outputs of both ($\boldsymbol{O}_1$and $\boldsymbol{O}_2$) are flattened separately and then concatenated to get a vector, followed by a FCL to get $\boldsymbol{h}$.

2. After calculating by the Softmax function, the weight of the upper and lower are h_1 and h_2.

3. Both outputs are multiplied by their respective weights and finally packed to output the final feature vector. The process is shown below:

$$\boldsymbol{h} = \boldsymbol{W} \cdot \text{Concat}(\boldsymbol{O}_1, \boldsymbol{O}_2) + \boldsymbol{b} \tag{16}$$

$$[h_1, h_2] = \text{Softmax}\,(\boldsymbol{h}) \tag{17}$$

$$\boldsymbol{y} = \text{Concat}(\boldsymbol{O}_1.h_1, \boldsymbol{O}_2, h_2) \tag{18}$$

V. EXPERIMENTAL ANALYSIS

A. EXPERIMENTAL DATA AND ENVIRONMENT

A combat simulation system provides the experimental data used in this paper. We construct the following scenario: in an early warning air defense combat, the enemy is conducting operations against our essential entity, which can receive air target movement status information from the early warning intelligence fusion center, but the information contains noise and outliers. The intentions of the different enemy targets vary, but they are all directed at a single important target. Note that our proposed model is primarily used to identify the intentions of air targets and to provide a reliable basis and aid for the following command, but it is not used for subsequent decision-making. In the experimental data, the time series characteristics data are derived from the backend of the simulation system, and they contain Gaussian white noise and a 15∼20% proportion of outliers, and the measurement noises are $\sigma_v^2 = (5m/s)^2$, $\sigma_d^2 = (1000\,|\,m)^2$, $\sigma_h^2 = (100\,|\,m)^2$, $\sigma_{\theta_1}^2 = \sigma_{\theta_2}^2 = (1\,°)^2$, and $\sigma_a^2 = (1m/s^2)^2$, respectively. The index of the location where the outlier occurs is random; the labels are given by early warning intelligence and air warfare experts. In the alternative data, 1000 samples were selected for each class of intention, the data reception frequency is 10 Hz, and the sample length is roughly between [1000, 3000]; 80% of the sample set is used to build the train set, and the remainder is used to build the test set. A sliding window with a step of two was used to slice the data of two sets for offline and online phases. The model input is a matrix of 6×s, where s is the sliding window length, which was determined through experiments.

B. EVALUATION METRICS

The model is trained using the train set, and the test set is used for model performance evaluation. Accuracy, Precision, Recall, F1 Score, and Loss are used to assess the model's performance.

1) ACCURACY

Accuracy is the ratio of the number of samples correctly predicted to the total number of samples in the test set.

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \tag{19}$$

where TP, TN, FP, and FN represent the number of samples whose true labels are positive and the classification results are positive, the number of samples whose true labels are negative and the classification results are negative, the number of samples whose true labels are negative but the classification results are positive, and the number of samples whose true labels are positive but the classification results are negative.

2) F1 SCORE

The F1 Score is the summed average of Precision and Recall. Specifically, the expressions for Precision, Recall and F1 score are as follows:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{20}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{21}$$

$$\mathrm{F1Score} = \frac{2 \times \mathrm{Precison} \times \mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \tag{22}$$

For the multiclassification problem, we use the composite F1 Score as an indicator with the following expression:

$$\text{Composite F1 Score} = \frac{1}{K}\sum\nolimits_{k=1}^{K} \mathrm{F1}_k \tag{23}$$

where K denotes the number of categories of intention, and $\mathrm{F1}_k$ denotes the F1Score corresponding to the k-th intention.

3) LOSS

Loss is the cross-entropy loss of the model on the test set, denoted as L:

$$L(y,\hat{y}) = -\frac{1}{K}\sum\nolimits_{k=1}^{K}(y\ln\hat{y} + (1-y)\ln(1-\hat{y})) \tag{24}$$

where K denotes the number of samples in the test set, y denotes the label, and $\hat{y}$ denotes the recognition result.

TABLE 3. The recognition performance of different sliding window lengths.

Sliding window length s	Accuracy (%)	Time (ms)
8	81.82	10.11
16	82.49	12.70
32	87.17	13.61
64	92.91	14.13
96	95.27	14.41
128	96.90	14.70
160	96.83	15.23

C. SLIDING WINDOW LENGTH DETERMINATION

Sliding window slicing of the raw time series data is required in both the offline training & learning and online recognition phases to increase the information used. The sliding window's length directly determines the model input's size and significantly impacts training and recognition. Too small a sliding window will result in utilizing too little information, making it difficult for the model to adequately extract implicitly discriminative information and thus fail to learn the mapping relationship between air target features and intention; too large a sliding window length will result in a significant increase in the computational cost of the model and increase the training time of the model during the offline phase and recognition time during the online phase. It is, therefore, necessary to determine the appropriate length of the sliding window.

Table 3 shows that the model's accuracy gradually improves, and the time used gradually increases as the s increases. And from Figure 7, we can see that the recognition accuracy fluctuates widely, and the model has difficulty converging when s is set too small (e.g. s= 8,s= 16). The convergence is significantly improved when $s > 64$. And when s is set to 128, the recognition accuracy reaches 96.90%, after which increasing the s value does not result in an increase in accuracy. However, instead, the time continues to increase. Considering recognition accuracy and time, the sliding window length s is set to 128, that is, the input size of the model is 6 × 128.

D. PARAMETER DETERMINATION

Hyperparameters have a significant impact on the recognition performance of deep learning algorithms. Therefore, it is necessary to set up comparison experiments to determine the values of some hyperparameters to improve the performance of intention recognition. Here the Adagrad optimizer is selected for the model, and the number of iterations EPOCH, batch size BS, learning rate LR, dropout value DP, Etc., are the primary hyperparameters to consider [51]. Based on experience in designing deep learning networks, the alternative hyperparameters set is set as EPOCH = {20,30,40,50}, BS = {32,64,96,128}, LR = {0.001,0.002,0.003} and

TABLE 4. Accuracy with different hyperparameters.

EPOCH	BS=64			
	DP	0.2	0.4	0.6
		LR=0.001/0.002/0.003	LR=0.001/0.002/0.003	LR=0.001/0.002/0.003
20		95.73/96.81/95.02	95.30/96.90/87.68	91.91/96.06/87.68
30		95.78/96.80/96.00	95.46/96.90/88.38	91.91/96.06/87.68
40		95.82/96.80/96.23	95.39/96.89/93.21	94.83/96.31/93.21
50		95.73/96.80/96.25	95.39/**96.90**/94.04	94.81/96.34/94.04
	BS=96			
	DP	0.2	0.4	0.6
		LR=0.001/0.002/0.003	LR=0.001/0.002/0.003	LR=0.001/0.002/0.003
20		95.42/96.65/92.06	95.49/96.03/83.20	95.07/95.91/78.88
30		95.45/96.67/93.54	95.47/96.60/92.74	95.17/96.37/86.61
40		95.46/96.63/94.91	95.51/96.60/95.57	95.24/96.37/90.66
50		95.46/96.66/94.84	95.39/96.56/96.18	95.18/96.40/93.60
	BS=128			
	DP	0.2	0.4	0.6
		LR=0.001/0.002/0.003	LR=0.001/0.002/0.003	LR=0.001/0.002/0.003
20		95.36/96.12/92.06	95.66/96.10/75.90	94.56/93.75/78.83
30		95.64/96.18/93.54	95.81/96.66/77.60	94.97/95.85/82.50
40		95.60/96.18/94.91	95.71/96.64/65.37	95.09/95.87/83.51
50		95.54/96.22/94.84	95.71/96.63/81.55	95.07/95.86/92.29

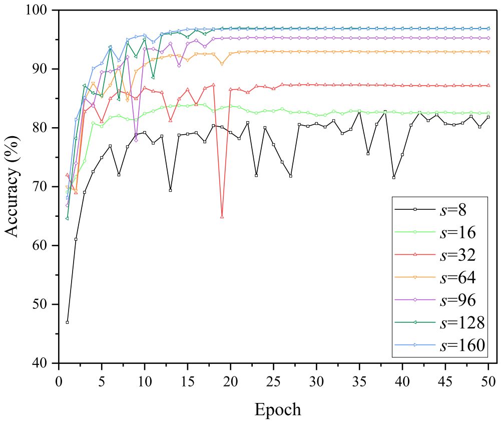


FIGURE 7. The recognition accuracy with different sliding window lengths.

TABLE 5. Hyperparameter and network parameter of WSPTCTE.

hyperparameter or network parameter	symbol	value
epoch	EPOCH	50
batch size	BS	64
Optimizer		Adagrad
learning rate	LR	0.002
dropout probability value	DP	0.4
embedding dimension & sub-layer output dimension	d_{model}	512
hidden size in feed forward network	d_{hidden}	1024
the number of attention heads	h	8
query and key dimensions	d_k, d_k	8, 8
value dimension	d_v	8

$DP = \{0.2,0.4,0.6\}$. Moreover, the accuracy in the test set is used to evaluate the performance.

The results are shown in Table 4. From Table 4, the highest recognition accuracy of the model on the test set is 96.90% when the hyperparameters are BS=64, LR=0.002, DP=0.4, and EPOCH=50. Therefore, epoch number EPOCH, batch size BS, learning rate LR, and dropout probability DP are finally set to 50, 64, 0.002, and 0.4.

E. EXPERIMENTS AND ANALYSIS

1) RECOGNITION PERFORMANCE ANALYSIS

After above experiments, the final hyperparameters and parameters are determined, shown in the Table 5. The experimental results of the recognition model with the parameters and hyperparameters above are shown in Figure 8.

In Figure 8, we can see that the loss decreases sharply until 30 epochs, but after 30 epochs, the decrease is no longer significant, which indicates that the model converges in about 30 epochs. After the model is trained, the accuracy of the train set reaches 100%, and the loss value is approximately 0.0015; the accuracy of the test set reaches 96.9%, and the corresponding loss value is approximately 0.07. In addition, we analyze the recognition performance of the model for samples with different intention. The confusion matrix of the test set is shown in Figure 9, where different colors of the right heat map scale indicate different levels of recognition accuracy. The darker the color means the higher the accuracy.

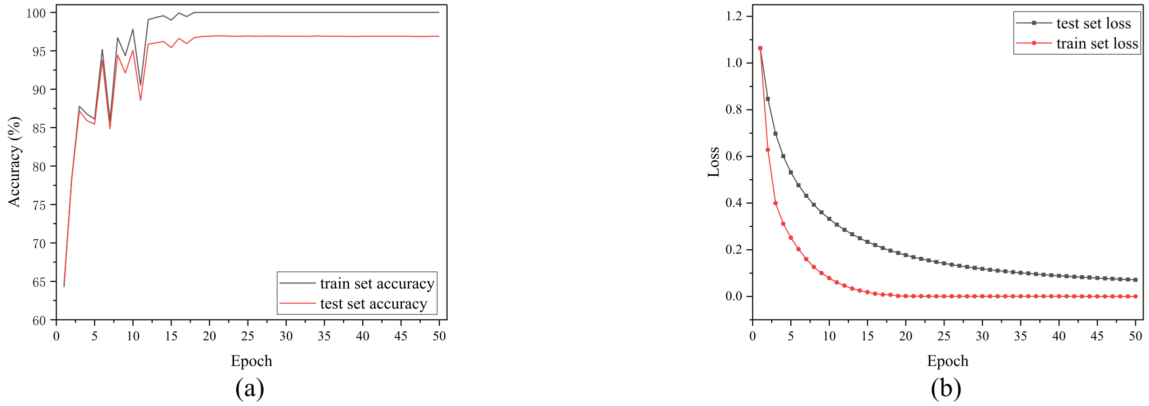


FIGURE 8. (a) The accuracy of WSPTCTE-IR; (b) The loss value of WSPTCTE-IR.

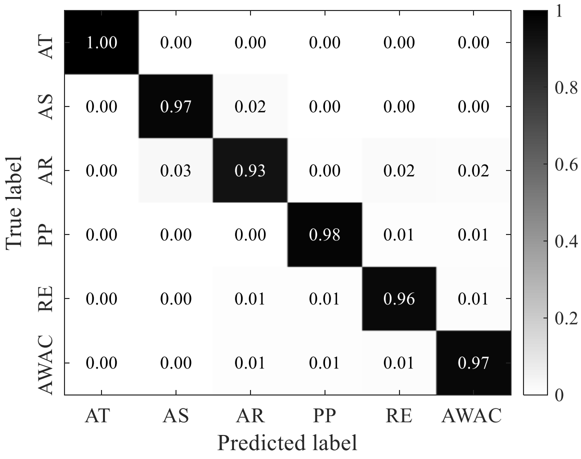


FIGURE 9. Confusion matrix of the WSPTCTE-IR.

And the AT, AS, AR, PP, RE, and AWAC represent attack, anti-submarine, aerial refueling, police patrol, retreat, and airborne warning and control.

Figure 9 shows that the model we propose has a high recognition accuracy for all six intention samples. Moreover, the recognition accuracy rate of attack intention is 100%, the highest among them, which means that the samples of attack intention can all be identified successfully. This result is mainly due to the special maneuvering status of targets with attack intention: they often take low-altitude and high-altitude penetration, resulting in a sharp rise or fall in altitude and a sharp reduction or increase in distance in a short time. Targets with other intentions are mainly unable and will not carry out such actions. Of the six intentions, the model has the lowest recognition accuracy of about 93% for the samples with aerial refueling intention. The results showed that the model would misclassify 3% of aerial refueling samples as anti-submarine and 2% of the anti-submarine samples as aerial refueling, which is due to the significant near-periodic variation in features such as heading angle, azimuth, and distance for both the anti-submarine and aerial refueling intention samples, which to some extent, leads to partially incorrect identification results.

Sample runs under each intention category (except attack intention) are also performed to intuitively show the performance of the proposed model. Figure 10 presents the predicted intention during the recognition process.

For the selected sample with anti-submarine intention, the recognition model outputs correct results after about 1.2 s, after which the output values fluctuate. After approximately 3 s, the output values remained mainly correct and stable. And the model outputs correct and stable identification results after about 2.9 s for aerial refueling intention. For the sample with police patrol intention, the model outputs correct results after about 0.6 s, and the recognition result remains mainly stable and correct. For the retreat intention sample, note that after about 0.2 s, the intention has been identified correctly first, but the identification result does not remain stable until 5.8 s. Finally, the model outputs correct and stable identification results after about 1.8 s for AWAC intention. The above results show that the recognition results stabilize over time, and the rapidness and stability of the established model are proved.

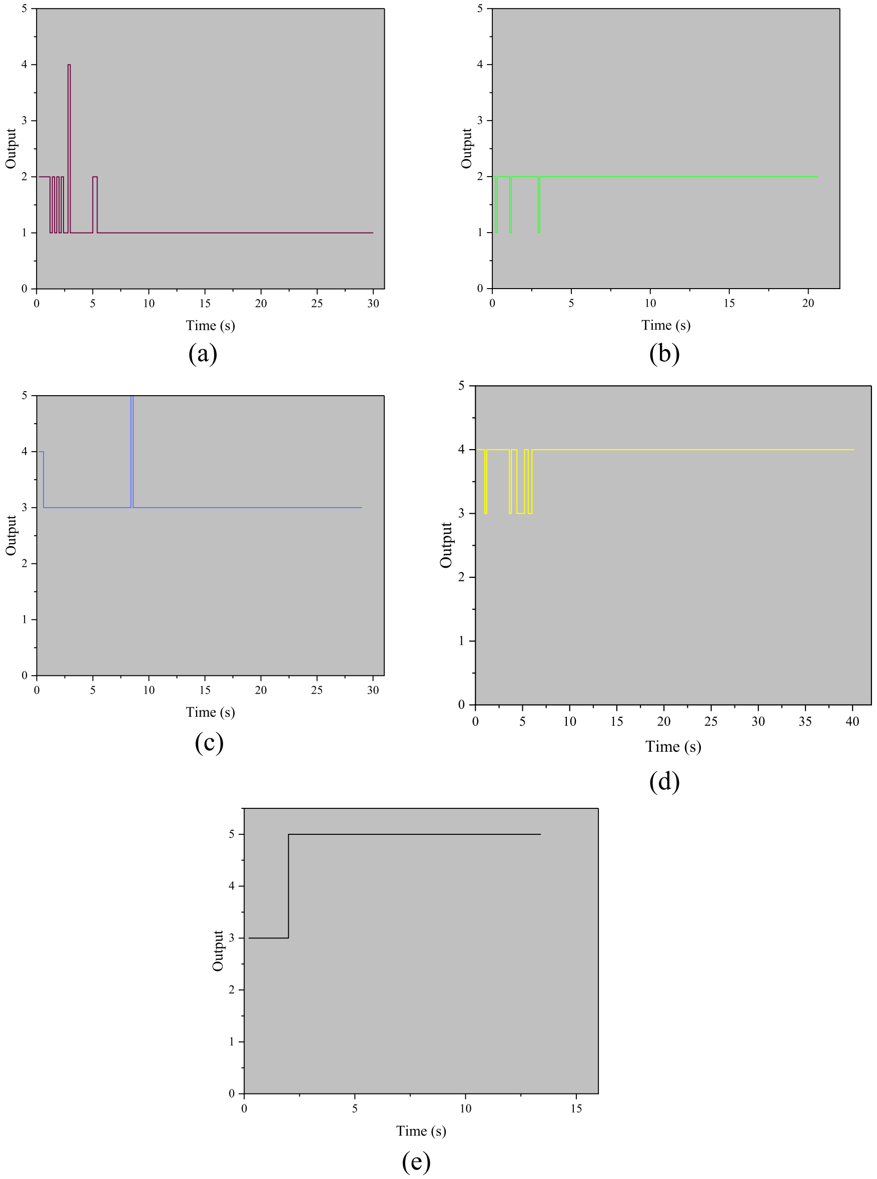
2) ROBUSTNESS TESTS

We then explore the recognition performance of the model for different noise levels and proportions of outliers. The noise level is defined as

$$M'^{(t)} = [d, v, a, \theta_1, \theta_2, \Delta H]^T + v(t)\, NL + u(t) \qquad (25)$$

where *NL* represents the noise level. When we explore the impact of noise levels on performance, it should be noted that the percentage of outliers is kept at 15∼20%. Similarly, when we explore the impact of the proportion of outliers on performance, the noise level remains 1.

The results of robustness tests are presented in Table 6 and Table 7. In Table 6, we can see that as the noise level increases, the accuracy of intention recognition becomes less accurate. However, at a noise level of 6, the recognition accuracy still exceeds 90%.



(a)

(b)

(c)

(d)

(e)

FIGURE 10. Sample run of different intention types (except attack intention). The output stepped line in (a), (b), (c), (d) and (e) represent the recognition process for the selected sample with AS, AR, PP, RE and AWAC intention. The output is coded in the same way as in Figure 3.

TABLE 6. The recognition accuracy on data with different noise levels.

Noise level	Accuracy (%)
1	96.90
2	93.70
4	93.16
6	90.41
8	88.39

TABLE 7. The recognition accuracy on data with different outlier proportion.

Outlier proportion (%)	Accuracy (%)
[15,20]	96.90
[20,25]	93.59
[25,30]	92.25
[30,35]	91.57

Table 7 shows the recognition accuracy concerning the proportion of outliers. The trend of the accuracy is similar to that of Table 6, i.e., as the ratio of outliers increases, the accuracy of recognition decays. Note that the accuracy of the recognition model still exceeds 90% when the outliers proportion reaches 30%. The identification results can still provide a useful reference for the commander's decision-making.

3) ABLATION EXPERIMENTS

Ablation experiments are conducted on the same dataset to verify the WSPTCTE-IR's effectiveness. X-Mask indicates

TABLE 8. Results of ablation experiments.

Model		Assembly				Accuracy (%)	Composite F1
Reference Number	Name	TTE	Masked MHA	CTE	WSU		
1	TTE	√				93.21	92.92
2	TTE-Mask	√	√			91.61	90.86
3	CTE			√		96.29	96.14
4	PTCTE	√		√		95.50	95.33
5	PTCTE-Mask	√	√	√		95.16	94.94
6	WSPTCTE-IR	√		√	√	**96.90**	**96.76**
7	WSPTCTE-MASK-IR	√	√	√	√	96.09	95.90

TABLE 9. The precision of different models.

Model		AT	AS	AR	PP	RE	AWAC
Reference Number	Name						
1	TTE	100	89.82	89.13	**98.74**	87.55	95.85
2	TTE-Mask	100	97.67	86.65	93.23	88.83	77.85
3	CTE	100	97.05	91.51	97.76	95.19	95.58
4	PTCTE	100	95.64	92.41	96.16	93.71	93.91
5	PTCTE-Mask	100	96.08	89.99	96.79	92.8	94.34
6	WSPTCTE-IR	**100**	**97.36**	**92.61**	97.84	**96.47**	**96.63**
7	WSPTCTE-MASK-IR	100	97.12	91.43	97.41	94.98	94.71

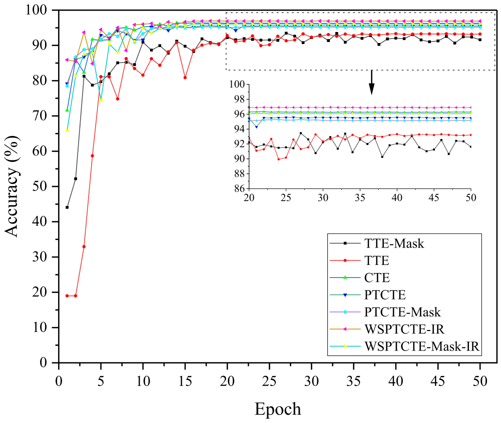


FIGURE 11. The recognition accuracy of models with different assembly combinations.

the use of masked MHA instead of MHA [48]. The setting and performance are presented in Table 8. Furthermore, the curves of recognition accuracy of models for different assembly combinations are shown in Figure 11.

From Table 8, we can see that the recognition accuracy of WSPTCTE-IR (Model 6) is optimal. Experimental results of TTE (Model 1) and TTE-Mask (Model 2) show that introducing Masked MHA decreases recognition performance. This is because the masked MHA reduces the data utilized on the classification task. In addition, directly concatenating the outputs of TTE and CTE (Model 3) can cause a decline in recognition performance. Specifically, the recognition accuracy of the parallel time-channel Transformer Encoder (PTCTE, Model 4) slips by approximately 0.8% compared to the CTE. In Figure 11, the seven models generally improve recognition accuracy as the training epochs increase, with WSPTCTE-IR consistently outperforming the other six models after 20 epochs. The results show that introducing WSU can effectively improve recognition accuracy. Compared to TTE, CTE, and PTCTE, the recognition accuracy is improved by 3.7%, 0.6%, and 1.4%, respectively.

Then the precision, recall, and F1 score are utilized to assess the recognition accuracy of the seven models, as shown in Table 9, Table 10, and Table 11. WSPTCTE-IR has the highest precision value, recall value, and F1 score for almost every intention. Consistent with the previous analysis, simply concatenating the TTE and CTE is hard to bring performance improvement, while introducing WSU can effectively learn the weights beneficial for classification, resulting in improved recognition performance. Furthermore, comparing the six types of intentions, the precision, recall, and F1 Score of the aerial refueling intention is lowest for most models because the feature of aerial refueling intention has the cyclical character that exists in some other intentions. The highest recall, precision, and F1 score are obtained for attack intention because the maneuvers and tactical actions of attack intention are apparent, and the model can learn its characters better.

4) COMPARISON EXPERIMENTS

Since no public dataset exists for air target intention recognition, we used other air target intention recognition methods from the references to conduct comparison experiments. In addition, due to the difficulty of defining the parameters of the model-based methods, some data-based methods are used for comparison in this paper.

The method used are XGboost [40], SVM [37], Random Forrest (RF) [38], FCN [43], CNN [43], LSTM [43], BiGRU-ATTENTION [52], CNN-BiLSTM-ATTENTION [53]. Under the same sliding window length, intention feature and space, the methods are trained, and the final recognition

TABLE 10. The recall of different models.

Model		AT	AS	AR	PP	RE	AWAC
Reference Number	Name						
1	TTE	100	**98.65**	90.45	83.54	90.66	92.53
2	TTE-Mask	100	93.14	94.83	80.62	86.59	93.87
3	CTE	100	96.01	94.40	96.81	94.96	94.51
4	PTCTE	100	95.71	90.36	97.05	94.42	**94.70**
5	PTCTE-Mask	100	95.64	91.92	96.22	93.47	92.09
6	WSPTCTE-IR	**100**	96.87	**95.51**	**97.66**	**95.72**	94.58
7	WSPTCTE-MASK-IR	99.98	96.26	94.53	96.81	94.31	93.34

TABLE 11. The f1-score of different models.

Model		AT	AS	AR	PP	RE	AWAC
Reference Number	Name						
1	TTE	1.0000	0.8979	0.9403	0.9051	0.8908	0.9416
2	TTE-Mask	1.0000	0.9535	0.9056	0.8647	0.8770	0.8511
3	CTE	1.0000	0.9652	0.9293	0.9728	0.9508	0.9504
4	PTCTE	1.0000	0.9568	0.9137	0.9660	0.9406	0.9430
5	PTCTE-Mask	1.0000	0.9586	0.9095	0.9650	0.9313	0.9320
6	WSPTCTE-IR	**1.0000**	**0.9711**	**0.9404**	**0.9775**	**0.9609**	**0.9559**
7	WSPTCTE-MASK-IR	0.9999	0.9669	0.9295	0.9711	0.9464	0.9402

TABLE 12. The recognition performance comparison of the method our proposed and works from other papers.

Model	Accuracy (%)	Composite F1 score
XGboost [40]	92.33	0.9206
SVM [37]	80.20	0.7901
RF [38]	81.75	0.7912
FCN [43]	85.25	0.8319
CNN [43]	89.84	0.8934
LSTM [43]	92.67	0.9205
BiGRU-ATTENTION [52]	95.11	0.9491
CNN-BiLSTM-ATTENTION [53]	91.17	0.9058
WSPTCTE-IR	**96.90**	**0.9676**

performance are obtained. The comparison experiments results are shown in Table 12.

From results in Table 12, we can see that the recognition accuracy and F1 Score of WSPTCTE-IR are higher than those of several air target intention recognition approaches. Specifically, the recognition accuracy is 96.9% and the composite F1 score is 0.9676.

The comparison shows that deep-learning-based methods outperform most machine-learning-classifier-based methods. In machine-learning-classifier-based methods, XGboost outperforms others in accuracy and Composite F1 score, which indicates that tree boosting strategy allows for more effective air target intention recognition. Among deep-learning-based methods, RNN-based methods outperform CNN-based methods in general. In addition, most of the methods that uses the attention mechanism improves recognition accuracy. Finally, it is worth noting that more complex network structures do not necessarily lead to performance gains, e.g., the recognition accuracy of CNN-BiLSTM-ATTENTION is 91.17%, which is lower than the recognition accuracy of LSTM and BiGRU-ATTENTION.

VI. CONCLUSION

To enable commanders and operators to effectively analyze the battlefield situation and improve the rationality of decision-making when the acquired information contains noise and outliers, a robust air target intention recognition model based on the WSPTCTE is proposed in this paper. The input to the model is enemy target state information containing noise and outliers, and the output is the intention label. The model extracts step-wise and channel-wise correlations by building a Transformer Encoder model on the time and channel axes. It also introduces WSU to automatically learn the weights of two parallel branches to avoid possible performance decay from directly concatenating them. Experimental results show that the method has higher recognition accuracy and F1 score than other methods. In addition, the performance is explored when the noise level and the ratio of outliers increase, and the results demonstrate the robustness of the model. Finally, ablation experiments verify the effectiveness of the model.

Furthermore, the analysis of the experimental results shows that the model's recognition process aligns with the general commanders' situational awareness thinking. In future, we plan to conduct further research on air target intention recognition in combat scenarios with more confrontation, more detailed intention granularity and a wider variety of targets to further increase the generality and robustness of the proposed model.

REFERENCES

[1] H. A. Kautz and J. F. Allen, "Generalized plan recognition," in *Proc. 4th AAAI Nat. Conf. Artif. Intell.*, Philadelphia, PA, USA, 1986, pp. 32–37.

[2] M. Zhao, H. Gao, W. Wang, and J. Qu, "Research on human-computer interaction intention recognition based on EEG and eye movement," *IEEE Access*, vol. 8, pp. 145824–145832, 2020.

[3] D. Wei, L. Chen, L. Zhao, H. Zhou, and B. Huang, "A vision-based measure of environmental effects on inferring human intention during human robot interaction," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4246–4256, Mar. 2022.

[4] J. Ma, X. Guo, and X. Zhao, "Identifying purchase intention through deep learning: Analyzing the Q&D text of an e-commerce platform," *Ann. Oper. Res.*, pp. 1–20, Jul. 2022. [Online]. Available: https://doi.org/10.1007/s10479-022-04834-w

[5] M. Yang, D. Wang, S. Feng, and Y. Zhang, "An empirical study on learning based methods for user consumption intention classification," in *Natural Language Processing and Chinese Computing*, X. Huang, J. Jiang, D. Zhao, Y. Feng, and Y. Hong, Eds. Cham, Switzerland: Springer, 2018, pp. 910–918.

[6] X. Ding, T. Liu, J. Duan, and J.-Y. Nie, "Mining user consumption intention from social media using domain adaptive convolutional neural network," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2389–2395.

[7] M. Che, Y. D. Wong, K. M. Lum, and M. C. Rojas Lopez, "Users' behavioral intention and their behavior: Before-and-after study of 'keep left' markings on shared footpaths," *Int. J. Sustain. Transp.*, vol. 17, no. 3, pp. 219–227, Mar. 2023.

[8] R. Quan, L. Zhu, Y. Wu, and Y. Yang, "Holistic LSTM for pedestrian trajectory prediction," *IEEE Trans. Image Process.*, vol. 30, pp. 3229–3239, 2021.

[9] R. Q. Mínguez, I. P. Alonso, D. Fernández-Llorca, and M. Á. Sotelo, "Pedestrian path, pose, and intention prediction through Gaussian process dynamical models and pedestrian activity recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1803–1814, May 2019.

[10] H. Huang, Z. Zeng, D. Yao, X. Pei, and Y. Zhang, "Spatial-temporal ConvLSTM for vehicle driving intention prediction," *Tsinghua Sci. Technol.*, vol. 27, no. 3, pp. 599–609, Jun. 2022.

[11] Z. Wu, K. Liang, D. Liu, and Z. Zhao, "Driver lane change intention recognition based on attention enhanced residual-MBi-LSTM network," *IEEE Access*, vol. 10, pp. 58050–58061, 2022.

[12] Y. Xia, Z. Qu, Z. Sun, and Z. Li, "A human-like model to understand surrounding vehicles' lane changing intentions for autonomous driving," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4178–4189, May 2021.

[13] L. Chen, X. Liang, Y. Feng, L. Zhang, J. Yang, and Z. Liu, "Online intention recognition with incomplete information based on a weighted contrastive predictive coding model in wargame," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7515–7528, Oct. 2023, doi: 10.1109/TNNLS.2022.3144171.

[14] C. D. Wickens, "Situation awareness: Its applications value and its fuzzy dichotomies," *J. Cognit. Eng. Decis. Making*, vol. 9, no. 1, pp. 90–94, Mar. 2015.

[15] A. Munir, A. Aved, and E. Blasch, "Situational awareness: Techniques, challenges, and prospects," *AI*, vol. 3, no. 1, pp. 55–77, Mar. 2022.

[16] H. Xu, J. Zhao, L. Chen, W. Tan, and H. Zhang, "A review of methods of battlefield target combat intention recognition," in *Proc. Int. Conf. Auton. Unmanned Syst.* (Lecture Notes in Electrical Engineering), W. Fu, M. Gu, and Y. Niu, Eds. Singapore: Springer, 2023, pp. 3686–3696.

[17] J. Johnson, "Artificial intelligence and the future of warfare: The USA, China, and strategic stability," in *Artificial Intelligence and the Future of Warfare*. Manchester, U.K.: Manchester Univ. Press, 2021.

[18] X. Liang, L. Liu, J. Zhang, and S. Li, "Aviation swarm and intelligent air combat," in *Proc. IEEE CSAA Guid., Navigat. Control Conf. (CGNCC)*, Aug. 2018, pp. 1–6.

[19] R. Alford, H. Borck, and J. Karneeb, "Active behavior recognition in beyond visual range air combat," in *Proc. 3rd Annu. Conf. Adv. Cognit. Syst.*, 2015, pp. 1–10.

[20] F. A. Van-Horenbeke and A. Peer, "Activity, plan, and goal recognition: A review," *Frontiers Robot. AI*, vol. 8, May 2021, Art. no. 643010.

[21] A. Dahlbom, "A comparison of two approaches for situation detection in an air-to-air combat scenario," in *Modeling Decisions for Artificial Intelligence* (Lecture Notes in Computer Science), V. Torra, Y. Narukawa, G. Navarro-Arribas, and D. Megias, Eds. Berlin, Germany: Springer, 2013, pp. 70–81.

[22] H. Borck, J. Karneeb, R. Alford, and D. W. Aha, "Case-based behavior recognition in beyond visual range air combat," in *Proc. FLAIRS Conf.*, 2015, pp. 379–384.

[23] Q. Xiao, Y. Liu, X. Deng, and W. Jiang, "A robust target intention recognition method based on dynamic Bayesian network," in *Proc. 33rd Chin. Control Decis. Conf. (CCDC)*, May 2021, pp. 6846–6851.

[24] Z. Zhang, Y. Qu, and H. Liu, "Air target intention recognition based on further clustering and sample expansion," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 3565–3569.

[25] W. Xiang, X. Li, Z. He, C. Su, W. Cheng, C. Lu, and S. Yang, "Intention estimation of adversarial spatial target based on fuzzy inference," *Intell. Autom. Soft Comput.*, vol. 35, no. 3, pp. 3627–3639, 2023.

[26] X. Xia, "The study of target intent assessment method based on the template-matching," M.S. thesis, Graduate School, Nat. Univ. Defense Technol., Changsha, China, 2006.

[27] H. Leng, X. Wu, J. Hu, and Y. Yong, "Study on sequential recognition technique of marine targets' tactical intentions," *Syst. Eng. Electron.*, vol. 3, pp. 462–465, Jan. 2008.

[28] M. Li, X. Feng, and W. Zhang, "Template-based inference model and algorithm for simulation assessment in information fusion," *Fire Control Command Control*, vol. 35, no. 6, pp. 64–66, 2010.

[29] Y. Song, X. Zhang, and H. Guo, "Hierarchical inference frame and realization of air target tactical intention," *Inf. Command Control Syst. Simul. Technol.*, vol. 27, no. 5, pp. 63–66, 2005.

[30] Z. Wu and D. Li, "A model for aerial target attacking intention judgment based on reasoning and multi-attribute decision making," *Electron. Opt. Control*, vol. 17, no. 5, pp. 10–13, 2010.

[31] B. Li, P. Fan, and L. Tian, "Combat intention identification of target based on sequential three-way decision," *J. Shaanxi Normal Univ.*, vol. 50, no. 3, pp. 17–23, 2022.

[32] C. Yang, S. Song, and P. Fan, "A target intention recognition method based on cost-sensitive and multi-class three-branch decision," *J. Ordnance Equip. Eng.*, vol. 44, no. 2, pp. 132–136, 2023.

[33] S. Wang, G. Wang, Q. Fu, Y. Song, J. Liu, and S. He, "STABC-IR: An air target intention recognition method based on bidirectional gated recurrent unit and conditional random field with space-time attention mechanism," *Chin. J. Aeronaut.*, vol. 36, no. 3, pp. 316–334, Mar. 2023.

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

[35] J. Qing, G. Xiantai, J. Weidong, and W. Nanfang, "Intention recognition of aerial targets based on Bayesian optimization algorithm," in *Proc. 2nd IEEE Int. Conf. Intell. Transp. Eng. (ICITE)*, Sep. 2017, pp. 356–359.

[36] Y. Xu, S. Cheng, H. Zhang, and Z. Chen, "Air target combat intention identification based on IE-DSBN," in *Proc. Int. Workshop Electron. Commun. Artif. Intell. (IWECAI)*, Jun. 2020, pp. 36–40.

[37] M. Guanglei, Z. Runnan, W. Biao, Z. Mingzhe, W. Yu, and L. Xiao, "Target tactical intention recognition in multiaircraft cooperative air combat," *Int. J. Aerosp. Eng.*, vol. 2021, pp. 1–18, Nov. 2021.

[38] Z. Hu, H. Liu, S. Gong, and C. Peng, "Target intention recognition based on random forest," *Modern Electron. Technique*, vol. 45, no. 19, pp. 1–8, 2022.

[39] Z. Yang, Z.-X. Sun, H.-Y. Piao, J.-C. Huang, D.-Y. Zhou, and Z. Ren, "Online hierarchical recognition method for target tactical intention in beyond-visual-range air combat," *Defence Technol.*, vol. 18, no. 8, pp. 1349–1361, Aug. 2022.

[40] W. Lei and L. Shizhong, "Tactical intention recognition of aerial target based on XGBoost decision tree," *J. Meas. Sci. Instrum.*, vol. 9, no. 2, pp. 148–152, 2018.

[41] J. Xue, J. Zhu, J. Xiao, S. Tong, and L. Huang, "Panoramic convolutional long short-term memory networks for combat intension recognition of aerial targets," *IEEE Access*, vol. 8, pp. 183312–183323, 2020.

[42] F. Teng, Y. Song, and X. Guo, "Attention-TCN-BiGRU: An air target combat intention recognition model," *Mathematics*, vol. 9, no. 19, p. 2412, Sep. 2021.

[43] C. Qu, Z. Guo, S. Xia, and L. Zhu, "Intention recognition of aerial target based on deep learning," *Evol. Intell.*, pp. 1–9, May 2022. [Online]. Available: https://doi.org/10.1007/s12065-022-00728-9

[44] Y. Wang, J. Wang, S. Fan, and Y. Wang, "Quick intention identification of an enemy aerial target through information classification processing," *Aerosp. Sci. Technol.*, vol. 132, Jan. 2023, Art. no. 108005.

[45] X. Wang, Z. Yang, G. Zhan, J. Huang, S. Chai, and D. Zhou, "Tactical intention recognition method of air combat target based on BiLSTM network," in *Proc. IEEE Int. Conf. Unmanned Syst. (ICUS)*, Oct. 2022, pp. 63–67.

[46] T. Zhou, M. Chen, Y. Wang, J. He, and C. Yang, "Information entropy-based intention prediction of aerial targets under uncertain and incomplete information," *Entropy*, vol. 22, no. 3, p. 279, Feb. 2020.

[47] L. Carlson, D. Navalta, M. Nicolescu, M. Nicolescu, and G. Woodward, "Early classification of intent for maritime domains using multinomial hidden Markov models," *Frontiers Artif. Intell.*, vol. 4, Oct. 2021, Art. no. 702153.

[48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. 2017, pp. 1–11.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[50] J. Lei Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.

[51] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, no. 7, pp. 257–269, 2011.

[52] F. Teng, X. Guo, Y. Song, and G. Wang, "An air target tactical intention recognition model based on bidirectional GRU with attention mechanism," *IEEE Access*, vol. 9, pp. 169122–169134, 2021.

[53] X. Wang, Z. Lin, Y. Hu, and J. Liu, "Learning embedding features based on multisense-scaled attention architecture to improve the predictive performance of air combat intention recognition," *IEEE Access*, vol. 10, pp. 104923–104933, 2022.

ZIHAO SONG received the master's degree in electronics and communication engineering from the Air Force Early Warning Academy, Wuhan, China, in 2021, where he is currently pursuing the Ph.D. degree in information and communication engineering.

His research interests include situation awareness, signal recognition, and AI.

YAN ZHOU received the Ph.D. degree in control science and engineering from the 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.

WEI CHENG received the Ph.D. degree in control science and engineering from the Air Force Engineering University, Xi'an, China, in 2003.

He is currently an Associate Professor with the Air Force Early Warning Academy, Wuhan. His research interests include communication signals analysis, intelligence antennas, and radar-communication integration.

FUTAI LIANG received the master'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 in information and communication engineering.

His research interests include target detection, image recognition, and AI.

CHENHAO ZHANG received the master's degree in information science from the Air Force Early Warning Academy, Wuhan, China, in 2020, where he is currently pursuing the Ph.D. degree in information science.

His research interests include situation awareness, AI, and data mining.