Spatio-Temporal-Frequency Graph Attention Convolutional Network for Aircraft Recognition Based on Heterogeneous Radar Network

YUEXING PENG, YONGHUI LI, WENBO WANG, PENG CHENG, WEI XIANG

The University of Sydney, Sydney, NSW, Australia
La Trobe University, Melbourne, VIC, Australia
Beijing University of Posts and Telecommunications, Beijing, China

This article proposes a knowledge- and data-driven graph neural network-based collaboration learning model for reliable aircraft recognition in a heterogeneous radar network. The aircraft recognizability analysis shows that the semantic feature of an aircraft is motion patterns driven by the kinetic characteristics, and the grammatical features contained in the radar cross-section (RCS) signals present spatial-temporal-frequency (STF) diversity decided by both the electromagnetic radiation shape and motion pattern of the aircraft. Then, an spatio-temporal-frequency graph attention convolutional network (STFGACN) is developed to distill semantic features from the RCS signals received by the heterogeneous radar network. Extensive experiment results verify that the STFGACN outperforms the baseline methods in terms of detection accuracy, and ablation experiments are carried out to further show that the expansion of the information dimension can gain considerable benefits to perform robustly in the low signal-to-noise ratio region.

1. INTRODUCTION

Air travel is widely used in both military and civilian applications due to high mobility, long range, and flexible deployment. However, with the rapidly increasing capabilities of aircraft on defense penetration and low detectability in complex electromagnetic environments, airspace security and privacy becomes a very challenging task. As a countermeasure, aircraft recognition aims to classify target aircrafts by radar signals and has attracted worldwide attention in recent years, where intensive in-depth studies on radar-based aircraft recognition have become a research focus [1]–[16].

Radar based-aircraft recognition entails extracting characteristics from echo signals, and then classifying the target aircraft. The echo signal of the target aircraft contains rich semantic information, such as velocity, shape, size, and acceleration. Under the umbrella of low detectability technology, it now becomes very difficult to detect characteristic parameters, such as the electromagnetic radiation sharp information, from weak radar echo signals reliably. Therefore, it is imperative to develop effective techniques to extract useful information from radar echo signals reliably and accurately.

The commonly used radar echo signals can be roughly classified into three categories: 1) the radar cross-section (RCS) signal; 2) the high-resolution range profile (HRRP); and 3) the inverse synthetic aperture radar (ISAR) imaging. Although the HRRP [12]–[14] provides precise information with a higher spatial and temporal resolution than the RCS signal, it requires high-resolution hardware and ultrawideband radar. The ISAR [15], [16] mainly reflects the motion information, and is usually obtained from the airborne radar towards ground targets. Due to its easy availability and containing rich information of the target, the RCS signal is adopted in this article.

Single radar-based aircraft recognition methods can be roughly categorized into statistical feature-based methods and end-to-end deep learning ones. Statistical methods attempt to detect some key feature parameters, such as the Doppler frequency [4]–[6] or higher order cumulant [7], and then classify aircrafts using machine learning models. However, these methods require prior knowledge of the target aircraft, which may not be available, and their performance degrades severely in the low signal-to-noise ratio (SNR) region. Deep learning-based target recognition methods can automatically extract semantic features
from abundant samples with limited direct human intervention [10]–[12]. For example, the methods in [10] and [11] processed the low-probability-of-intercept radar signals for classification by convolutional neural network (CNN) and combining support vector machine (SVM) and autoencoder, respectively. In [12], a deep generative recurrent network was developed to improve classification accuracy. Although superior recognition accuracy is achieved, the data-driven deep learning methods depend heavily on both the quantity and quality of the manually labeled data, whose performance inevitably deteriorates at low SNRs. Moreover, the data-driven methods, lacking recognizable mechanics analysis and relying only on grammatical feature learning from RCS signals, are unable to fully exploit semantic features. Therefore, these methods cannot maximize the performance of aircraft recognition.

As a viable alternative, a homogeneous radar network [17] consists of several spatially distributed radars of the same carrier frequency. These radars work collaboratively to receive spatially anisotropic scatter response waves simultaneously, which can greatly enhance the capability of aircraft recognition by means of temporal–spatial information fusion. Aiming for the low SNR region, Xiang and Han [18] enhanced the ability of weak signal detection through a quantitative detection fusion method. Wang and Li [19] proposed a relaxed Chebyshev center covariance intersection algorithm to fuse local estimates when the cross-correlation of local estimation errors is unavailable. The distributed radar network multiframe detection method was proposed in [20], which outperforms the single radar multiframe detection method. The aforementioned methods extract the spatial–temporal information of the signals through signal-level fusion or decision fusion. However, there is still plenty of room for improving the detection performance, because these methods cannot extract grammatical and semantic information simultaneously. Besides, signal-level fusion methods require very high calibration accuracy between different radar information, so they are only suitable for homogeneous radar networks.

When the collaborative radars in the radar network have different carrier frequencies and bandwidths, the original spatio-temporal domain is transformed into the much broader temporal–spatial-frequency domain, resulting in a heterogeneous radar network that can further enhance the detection performance. In this article, we propose a framework to effectively distill and fuse temporal–spatial-frequency information in the heterogeneous radar network. Beyond signal-level fusion or decision fusion, we propose a model referred to as the spatio-temporal-frequency graph attention convolutional network (STFGACN), which is the first to exploit both grammatical and semantic features of the heterogeneous radar network. Based on the recognizability analysis of aircraft, the proposed STFGACN model maps the diverse grammatical features in the form of variation patterns of the RCS signal to the semantic feature determined by the unique movement pattern of the target aircraft. This greatly enhances the recognition performance in the low SNR region through fusing the diverse grammatical features in the temporal–spatial-frequency domain.

The main contributions of this article are highlighted as follows.

1) Based on the heterogeneous radar network, a network learning framework is developed to learn the mapping from semantic features determined by the unique movement pattern of the aircraft to the diverse grammatical features via temporal–spatial-frequency variation patterns of the RCS.

2) A novel STFGACN model is designed to distill semantic features and reliably detect aircraft by fusing diverse grammatical features in the form of information expansion from the temporal–spatial domain to the temporal–spatial-frequency domain. To the best of the authors’ knowledge, this model is the first work to distill temporal–spatial-frequency information by means of the graph neural network (GNN) for aircraft recognition.

3) Extensive experiments are carried out to verify the superiority of the proposed model on the reliability of aircraft recognition, especially in the low SNR region.

The rest of this article is organized as follows. Section II introduces the related work on the GNN. Section III formulates the aircraft recognition problem. Section IV introduces the proposed framework and designed model. The performance evaluation is shown in Section V. Finally, Section VI concludes this article.

Notation: $n$, $n$, $N$, and $\mathbb{N}$ represent the variable, vector, matrix, and set, respectively.

II. RELATED WORK

The proposed STFGACN model is based on the recent proposed GNN [21], which works in non-Euclidean space and is a powerful technique to handle the real-world data featured by arbitrary graphical structures. According to the type of information extracted and the structure of the model, GNN can be roughly divided into the following three categories.

1) Spatial–Temporal Propagation Models: Combining graph structure and time series features in the dynamic propagation process, GNN is proposed to model the spatial–temporal propagation of the disease state [22]. Based on the original GNN, many enhanced models are developed, e.g., the cross-location attention-based GNN [23] and spatio-temporal GNNs [24].

2) Spatial–Temporal Similarity-Based Models: Based on the similarity among subnets in the spatial–temporal domain, graph convolutional network (GCN)-based models are employed to distill spatial–temporal features for traffic forecasting [25], [26], weather forecasting [27], skeleton-based action recognition [28], and many other tasks [29], [30].
3) **Semantic Context Extraction Models:** By extracting syntax-based grammatical rules and context-based semantic of language, GNN models are designed to extract complex interactions between words or sentences in natural language processing [31]–[33].

The differences between our model and the aforementioned models are summarized as follows:

1) Our model distills the semantic features depending on the aircraft movement pattern, while spatial–temporal propagation models model the grammatical representation in the form of state diffusion.

2) Spatial–temporal similarity-based models extract semantic features based on the similarity of both the temporal pattern of time-series and the spatial structure (e.g., the road network in traffic forecasting, geography in weather forecasting, and the skeleton structure in skeleton recognition). By contrast, our model focuses on the identical semantic features determined by the movement pattern, rather than the spatial–temporal similarity. Moreover, the information dimension extension of the frequency domain represents a challenging task, which introduces a significantly increased complexity to handle the grammatical variation and diversity, and can present more information for aircraft recognition through feature distillation.

3) Semantic context extraction models utilize the grammatical rules and the semantic context of language, while our model takes advantage of the identical semantic features determined by the aircraft movement pattern.

### A. System and Signal Models

As shown in Fig. 1, a heterogeneous radar network consists of \( n \) spatially distributed radars, generally differing in carrier frequency, bandwidth, and pulse duration. When an aircraft passes, radars detect it independently, and then output heterogeneous RCS signals, which contain diverse grammatical features of the aircraft in the form of time-varying patterns.

It is well-known that the time-varying pattern of the RCS is determined by the characteristics of both the aircraft and radar, i.e., the movement pattern and electromagnetic shape of the aircraft, the carrier frequency, bandwidth, scan frequency, and pulse duration of the radar radio wave and the position relationship between the aircraft and radar [7].

### III. PRELIMINARY

To simplify the modeling of the RCS signals, two coordinates are used to describe the mass trajectory and micromotion of the aircraft, i.e., the aircraft-body coordinate frame \((X_a, Y_a, Z_a)\) for mass trajectory from the viewpoint of the radar and the radar coordinate frame \((X_e, Y_e, Z_e)\) for micromotion, as shown in Fig. 2. The received RCS signal at time \( t \) by the \( n \)th radar can be represented by

\[
m_n(t, f) = m_n(\theta_n(t, f), \phi_n(t, f))
\]

where \( f \) is the carrier frequency of the radar and \( \theta_n \) and \( \phi_n \) are the horizontal and vertical components of the radar line-of-sight angle, respectively, which are defined by

\[
\phi_n(t) = \rho \times \arccos \frac{\Delta y_n(t)}{\Delta x_n(t)} + \sqrt{1 - \rho^2} \times \sigma_{\phi} \times \text{norm}(t + 1)
\]

\[
\theta_n(t) = \rho \times \arccos \frac{\Delta z_n(t)}{\sqrt{\Delta x_n^2(t) + \Delta y_n^2(t) + \Delta z_n^2(t)}} + \sigma_{\theta} \sqrt{1 - \rho^2} \times \text{norm}(t + 1)
\]

where \( \text{norm}(t + 1) \) denotes the normal distribution over \([-1, 1]\), \( \rho \) is the correlation coefficient, \( \sigma_{\phi} \) and \( \sigma_{\theta} \) are the disturbance variances of the horizontal and vertical angles, respectively, and \( \Delta x_n(t), \Delta y_n(t), \text{ and } \Delta z_n(t) \) are the relative distances between the \( n \)th radar and the aircraft in the

![Heterogeneous radar network.](image1)

![Model of aircraft and radar sight angle.](image2)
the hidden state \( \eta \) and previous hidden state \( \eta \) through a linear combination of the following:

\[
\sigma \eta + \gamma V = \omega \eta
\]

where \( \eta \), \( \gamma \), and \( \sigma \) represent the yaw, pitch, and roll angles of the aircraft, respectively, and \((x_n, y_n, z_n)\) and \((x_n(t), y_n(t), z_n(t))\) represent the coordinates of the \( n \)th radar and the aircraft in the radar coordinate frame, respectively.

A cone-shaped aircraft [34] usually spins to maintain flight stability, which is one of the most important characteristic micromotions affecting the aspect angle. In other words, the time-varying pattern of RCS varies with the change of spatial observation position and the radar carrier frequency.

Aircraft recognition aims to classify target aircraft from the RCS signals received by the heterogeneous radar network. As shown in Fig. 3, the heterogeneous radar network with \( n \) radars can be viewed as an \( N \)-node undirected graph with weights \( G = (\mathcal{V}, \mathcal{E}, A) \), where \( \mathcal{V} \) consists of RCS vectors with \(|\mathcal{V}| = N\), \( \mathcal{E} \) is an edge set where an edge represents the connected nodes within the same subnet, and \( A \in \mathbb{R}^{N \times N} \) is the weight adjacency matrix whose elements represent the similarities between the RCS signals. We train a GNN model on \( G \) to learn a nonlinear complex function \( g(\cdot) \), which can map graph signal vectors to labels such that the target aircraft can be correctly classified.

IV. PROPOSED MODEL

In this section, we elaborate on the proposed STFGACN. As shown in Fig. 4, the structure of the proposed model consists of three layers: 1) a node layer; 2) a subnet layer; and 3) a global layer. In the node layer, each node independently extracts temporal features as a sample in the frequency domain. In the global layer, subnets are converged and semantic features are distilled from the spatial–temporal-frequency (STF) domain. The process in each layer is detailed one by one, and the overall network structure of the proposed model is shown in Fig. 5. In the node layer, gate recurrent unit (GRU) with attention mechanism (ATT) referred to as ATT-GRU is employed to extract temporal features at each node. In the subnet layer, the graph convolution network (GCN) and the ATT-GRU module are employed to extract temporal–spatial features from each homogeneous radar subnet. In the global layer, a decoder module is designed to distill semantic features from the temporal–spatial-frequency domain for accurate aircraft recognition.

A. Node Layer

As illustrated above, the time-varying pattern of RCS is determined by both the observed movement pattern of aircraft and the ranging pattern of radar, which are related to semantic and grammar features, respectively. The features in the RCS are the mass trajectory and micromotion features. The micromotion characteristics are related to high-frequency features, while the mass trajectory characteristics are related to low-frequency features. In order to extract both long- and short-term features from RCS, the GRU with the attention mechanism termed ATT-GRU is employed, whose structure is shown in Fig. 6.

In the proposed ATT-GRU module, the GRU unit [35] consists of update gates \( s_t \) and reset gates \( z_t \), for remembering or forgetting historic information. These two gates update the hidden state \( h_t \) through a linear combination of the candidate state \( c_t \) and previous hidden state \( h_{t-1} \) as follows:

\[
s_t = \sigma(W_x x_t + V_x h_{t-1} + b_t)
\]

\[
z_t = \sigma(W_x x_t + V_x h_{t-1} + b_t)
\]

\[
c_t = \tanh(W_x x_t + V_x (h_{t-1} \otimes z_t + b_t))
\]

\[
h_t = s_t \otimes c_t + h_{t-1} \otimes (1 - s_t)
\]
where $\sigma(\cdot)$ denotes the sigmoid function, $W_s, V_s, W_c, V_c, W_t$, and $V_t$ are the weight matrices, and $b_s, b_c, b_t$ are the bias vectors.

In order to extract long-term features, the attention mechanism is incorporated, which weighs the temporal features with different periods adaptively. The weight at the $i$th time step is calculated from hidden state $h_i$ by

$$u_i = \tanh(W_u h_i + b_u)$$ \hspace{1cm} (9)

where $W_u$ and $b_u$ are the weight matrices and bias, respectively. The weight $\alpha_i$ is obtained by normalizing the attention weights $u_i$ by the softmax function as follows:

$$\alpha_i = \text{softmax}(u_i) = \frac{\exp(u_i)}{\sum_{j=1}^{K} \exp(u_j)}$$ \hspace{1cm} (10)

where $K$ is the number of the hidden states.
Then, $h_i$ is weighted by $\alpha_i$ to yield the temporal features $o_t$ as follows:

$$o_t = \sum_{i=1}^{K} \alpha_i h_i.$$  \hspace{1cm} \text{(11)}

**B. Subnet Layer**

In the proposed model, a subnet is composed of nodes with the same carrier frequency, on which spatial–temporal convolution is applied to extract spatial–temporal features as sampled features in the frequency domain. The traditional GCN apply convolutions in spectral domains with graph Fourier transform, and the computational complexity is very high. To avoid this problem, 1stChebNet GCN [21] is adopted, so that we can build a deeper network by replacing the explicit parameterization via, e.g., the Chebyshev polynomials [30], with computation-reduced first-order polynomial, which realizes parameter sharing, fast and scalable graph calculation. The 1stChebNet GCN can be expressed as the following linear function:

$$g_G o \approx W_G (I_n + D^{-\frac{1}{2}}AD^{-\frac{1}{2}}) O$$  \hspace{1cm} \text{(12)}

where $g_G$ denotes the graph convolution operation based on the spectral graph convolution, $D \in \mathbb{R}^{N \times N}$ is a diagonal degree matrix, $W_G$ is the weighting coefficient, $A$ is the weight adjacency matrix, and $I_n$ is an identity matrix.

The simplified graph convolution operator can act as a temporal–spatial filter to distill local information by only taking neighboring nodes into consideration. By stacking the graph convolution filters, information from multistep neighboring nodes is converged, which achieves the same effect as K-localized convolution, but with a significantly reduced computational complexity.

Fig. 7 shows the proposed GCN module, which fuses spatial semantic features via the following tensor multiplication:

$$g_G o = W_G \hat{A} O$$  \hspace{1cm} \text{(13)}

where $\hat{A}=I_n + D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is the normalized adjacency matrix. Every node is updated by aggregating information of its neighborhood nodes. The fused spatial–temporal information is further distilled by the following ATT-GRU module, which shares the same structure as the one at the node layer.

**C. Global Layer**

In order to better extract semantic features determined by the movement pattern, a decoder module is designed to converge information from subnets, and then fuse temporal–spatial–frequency features by a CNN module, whose structure is depicted in Fig. 8. The decoder begins with a $5 \times 1$ convolution operation to extract semantic features from the feature-map yielded by the subnet layer. Then, layer normalization is performed to prevent overfitting before higher semantic feature extraction by a fully connected dense layer. The softmax activation function is used to generate aircraft classification results.

**V. NUMERICAL EXPERIMENTS**

**A. Experiment System**

In our experiments, $N = 9$ spatially distributed radars are divided into two subnets, each of which has different carrier frequencies. Specifically, the first subnet consists of five radars operating at 6.25 GHz, while the other four radars operate at 1.52 GHz to form the second subnet. Two types of aircrafts are considered in our experiments, which have the same electromagnetic (EM) scatter shape and the mass trajectory but different micromotions. The parameter settings of the aircrafts are listed in Table I, and the channel model is the additional white Gaussian noise (AWGN) channel, which is widely adopted when the pitch angle of the aircraft is large [4].

| Parameter          | type-1 aircraft | type-2 aircraft |
|--------------------|----------------|----------------|
| Mass trajectory    | 5km/s          |                |
| micro-motion       |                |                |
| $f_1$              | 0.64 Hz        | 2.75 Hz        |
| $f_2$              | 1.67 Hz        | 8.72 Hz        |
The RCS signals are obtained through EM simulation using the computer simulation technology (CST) simulation software [36], and the simulated RCS signals are shown in Fig. 9, from which we can easily distinguish these two types of aircrafts from the RCS without the AWGN. However, even when the noise is moderate, we cannot classify the aircrafts from the RCS signals directly.

B. Data Processing

One thousand RCS signal segments are simulated from each radar for each aircraft, and then 18,000 RCS signal segments are obtained as the sample set. An RCS signal segment lasts 10 s with 200 samples. The dataset is divided into the train, test, and validation sets with a ratio of 7:2:1.

The data is normalized by the maximum and minimum normalization method. According to the spatial distribution of the radars, the adjacency matrix \( A \) can be constructed, whose \((i, j)\)th element \( a_{ij} \) is given by

\[
a_{ij} = \begin{cases} 
0.3 + \frac{1}{d_{ij}}, & f_{c_i} = f_{c_j} \\
\frac{1}{d_{ij}}, & \text{otherwise}
\end{cases}
\]  

where \( f_{c_i} \) and \( f_{c_j} \) denote the operating frequencies of nodes \( i \) and \( j \), respectively, and \( d_{ij} \) indicates the distance between these two radars.

C. Model Parameter Setting and Environment

The experiments are implemented on a Linux server with CPU: Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40 GHz, and GPU: NVIDIA GeForce GTX 1080.

1) Performance Metrics: Four widely used performance metrics are employed: a) accuracy; b) precision; c) recall; and d) F1-score, which are defined as

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]  

\[
\text{Precision} = \frac{TP}{TP + FP}
\]  

\[
\text{Recall} = \frac{TP}{TP + FN}
\]  

\[
F1 = \frac{2 \cdot \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

where \( TP \) is the number of correctly recognized positive samples (the first aircraft type), \( TN \) is the number of correctly recognized negative samples (the second aircraft
Fig. 10. Accuracy of aircraft recognition versus the SNR for all concerned methods.

$F_P$ is the number of incorrectly recognized positive samples, and $F_N$ is the number of incorrectly recognized negative samples.

2) **Hyperparameters:** In temporal attention mechanisms, the input and hidden units of the GRU are set to 200 and 64, respectively. The size of the graph convolution kernel is set to 1 in the model with the first-order approximation. We train our model by the Adam optimizer [30] with an initial learning rate of 0.001. The batch size is 5 and the step decay rate is 0.5. The loss function is SigmoidCrossEntropyLoss. The training process stops after 100 epochs or ten nonimproving validation loss epochs.

3) **Baseline:** The STFGACN is compared with four baseline methods: a) FFT [37]; b) SVM [38]; c) GRU [35]; and d) STGCN [25].

These reference methods can be categorized into two groups: a) group 1 includes STGCN and STFGACN, which are based on the radar network; b) while group 2 consists of the other three models, which are single radar-based detection methods. For a fair comparison, the decision fusion strategy is employed for the group 2 methods. That is, a voting classifier is constructed where the nine single radar-based models act as element classifiers. These ensemble classifiers are denoted by FFT(9), SVM(9), and GRU(9) in the following.

D. Experimental Results

Fig. 10 shows the experimental results of all concerned model. The following observations can be made from Fig. 10.

1) The proposed STFGACN outperforms all reference models, and the performance gain increases with the decrease of the SNR, which verifies that the proposed STFGACN method can greatly enhance the reliability of aircraft recognition, especially in the low SNR region where traditional methods degrade severely. The performance gain of the proposed method stems from the STF domain feature extraction and semantic feature fusion, which is testified by the following experiments.

2) The GNN-based models are able to achieve better aircraft recognition than single radar-based counterparts, especially in the low SNR region. This result shows that the GNN-based methods can extract and fuse more useful information from multiple domains. Although the voting policy is deployed for the utilization of multidomain information, the group 2 methods have much lower efficiency in terms of fully exploiting the multidomain information, which is further validated by the following experiments when comparing the proposed method with the single radar-based methods without voting policy. The experimental results are shown in Fig. 11.

3) The results in Fig. 11 show that the performance enhancement is subtle when fusing the detection results from nine radars, and then yielding the final output via the voting policy for the single radar-based methods. Specifically, at SNR = −10 dB, the accuracy increases from 50.7, 50.3, and 51.1 to 51.6, 53.2, and 53% for FFT, SVM, and GRU, respectively. In other words, the decision-level fusion scheme for the single radar-based methods achieves no more than 3% performance gain, which indicates that the decision-level information fusion cannot achieve effective semantic feature distillation.

4) For the GNN-based methods, the proposed STFGACN can achieve about a 2-dB SNR gain over the STGCN model in the low SNR region. This SNR gain is attributed mainly to the temporal feature extraction by the ATT-GRU module, where a CNN is employed in the STGCN. As we know, ATT-GRU can better distill large-spanned temporal patterns than CNN, which is good at local feature distillation.

5) At higher SNR scenarios, all methods can reliably recognize aircrafts. That is, the accuracy of aircraft recognition is above 90% when the SNR is larger...
than 5 dB, and the accuracy approaches 100% when the SNR is above 10 dB. However, in practical applications, low SNRs are more representative scenarios. Given the threshold of recognition accuracy, an SNR gain means an increased radar detection range. For example, given the accuracy threshold of 75%, the proposed method can achieve a 4.7-dB SNR gain over the single radar-based methods, which is equivalent to a three-fold increase of the detection range.

E. Ablation Experiments

To understand the contribution to detection accuracy by each function module of the proposed model, ablation experiments are conducted, which can clarify the contribution of information dimension from STF domain. The compared variates of the proposed method include four models: 1) ATT-GRU; 2) STGCN; 3) STFGACN-1F; and 4) STFGACN-2F.

1) The ATT-GRU model is the simplified version when only one radar is used, and then it can only distill the temporal features of received signals. This model is also the enhanced version of GRU.
2) The STGCN model can extract temporal–spatial features by the CNN module, which can be regarded as the reduced version of the proposed model in the temporal–spatial domain by replacing ATT-GRU with CNN.
3) The STGACN-1F model denotes the radar network consisting of the same frequency, and it compares with the reference STGCN model to verify the temporal–spatial feature extraction capability.
4) The STGACN-2F model is the one used for the multiple-frequency radar network.

The experimental results are listed in Table II, where all models work at SNR = −10 dB in order to evaluate their robustness. We can observe the following from Table II.

1) The attention mechanism can offer 2% gain over the basic GRU model.
2) STFGACN-1F achieves 10.8% accuracy gain compared to ATT-GRU, which testifies that the extra feature in the spatial domain can greatly improve the accuracy.
3) STFGACN-1F achieves 2.8% accuracy gain over STGCN, which shows ATT-GRU can better extract features over large range.

4) STFGACN-2F can present 4.6% accuracy gain than STFGACN-1F, which shows that extra information can be distilled from the frequency domain.

It can be concluded from the abovementioned results that the spatial dimension information contributes to the accuracy improvement the most, and the semantic features from the temporal and spatial domains are also important for reliable recognition.

In order to quantify the performance gain attributed to each constituent module, we present the performance of detection accuracy with varying SNR settings.

1) Information Dimension Expansion–Temporal Dimension: Fig. 12 shows the accuracy versus the SNR for the GRU, ATT-GRU, and STFGACN-2F models. As can be seen from the figure, the attention mechanism can increase the accuracy by 3% in the low SNR region. In the experiments, since the feature frequencies of micromotion of the two types of aircrafts are: a) 0.64; and b) 1.67 Hz, respectively, the temporal periods are then about 1.56 and 0.6 s, which span about 31 and 12 samples, respectively. The attention mechanism can extract more long-period features from the RCS. It can be seen that when the threshold of detection accuracy is 75%, the ranging distance of the ATT-GRU is one and a half times of that of the GRU.

2) Information Dimension Expansion–Spatial Dimension: We present the experimental results of the ATT-GRU, STGCN, and STFGACN-1F methods over a range of SNRs in Fig. 13 to show the performance gain attributable to extra spatial information. The results in Fig. 13 show that the extracted spatial information can greatly enhance the detection accuracy, especially in low SNR region. STFGACN-1F has an accuracy improvement of 10% compared with ATT-GRU, when the SNR is below −5 dB. It can be concluded that the RCS signals received by the spatially dispersed radars have strong spatial correlation. By extracting and then fusing spatial features, the proposed method can greatly enhance the detection performance in
the low SNR region. When the detection accuracy threshold is 75%, the ranging distance of our method can be doubled compared to its single radar-based counterpart.

3) **Information Dimension Expansion–Frequency Dimension**: The performance comparison of the proposed STFGACN with the heterogeneous and homogeneous networks is shown in Fig. 14. By fusing features from multiple frequency domains, a 4.6% accuracy improvement is achieved. In terms of the ranging distance, the proposed method improves about 60% when the threshold of recognition accuracy is above 75%.

VI. CONCLUSION

In this article, the spatio-temporal-frequency graph attention convolutional network (STFGACN) model was proposed for aircraft recognition using the heterogeneous radar network. Spatially distributed radars with different operating at distinct frequencies constituted a heterogeneous radar network, which was modeled as an aircraft recognition graph. In the proposed model, the temporal, spatial, and frequency features were extracted and fused in the node, subnet, and global layers, respectively. In the node layer, the GRU with the attention mechanism was designed to extract temporal features at each node, and then the spatial features were extracted by the GCN module and then fused by the ATT-GRU module. In the global layer, the temporal–spatial-frequency information was fused and semantic features were distilled for aircraft classification. Extensive experimental results were presented to demonstrate that the proposed model can achieve notable improvements on both reliability and accuracy. Besides, ablation experiment results were carried out to show the expansion from the temporal domain to the temporal–spatial-frequency domain was of great significance for the improvement of reliability and accuracy, which further demonstrated the effectiveness of the proposed method in multidomain feature extraction and fusion.

REFERENCES

[1] A. Zyweck and R. E. Bogner
Radar target classification of commercial aircraft
*IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 598–606, Apr. 1996.

[2] B. Bhanu, D. E. Dudgeon, E. G. Zelnio, A. Rosenfeld, D. Casasent, and I. S. Reed
Guest editorial introduction to the special issue on automatic target detection and recognition
*IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 1–6, Jan. 1997.

[3] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang
Multi-view automatic target recognition using joint sparse representation
*IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 2481–2497, Jul. 2012.

[4] P. Molchanov, K. Egiazarian, J. Astola, A. Totsky, S. Leshchenko, and M. P. Jarabo-Amores
Classification of aircraft using micro-Doppler bicoherence-based features
*IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 1455–1467, Apr. 2014.

[5] L. Zhang, Y. Li, Y. Liu, M. Xing, and Z. Bao
Time-frequency characteristics based motion estimation and imaging for high speed spinning targets via narrowband waveforms
*Sci. China Inf. Sci.*, vol. 53, no. 8, pp. 1628–1640, 2010.

[6] Y. Yang, J. Lei, W. Zhang, and C. Lu
Target classification and pattern recognition using micro-Doppler radar signatures
In Proc. 7th ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw., Parallel/Distrib. Comput., 2006, pp. 213–217.

[7] J. Zhang, S. Hu, Q. Yang, and X. Fan
RCS statistical features and recognition model of air-floating corner reflector
*Syst. Eng. Electron.*, vol. 41, no. 4, pp. 780–786, 2019.

[8] L. Du, H. Liu, Z. Bao, and J. Zhang
Radar automatic target recognition using complex high-resolution range profiles
*IET Radar, Sonar, Navigation*, vol. 1, no. 1, pp. 18–26, 2007.

[9] D. Zhou, Y. Wang, X. Wang, X. Cheng, and J. Xiao
Radar emitter signal recognition based on deep restricted Boltzmann machine
*J. Nat. Univ. Defense Technol.*, vol. 38, no. 6, pp. 136–141, 2016.

[10] S.-H. Kong, M. Kim, L. M. Hoang, and E. Kim
Automatic LPI radar waveform recognition using CNN
*IEEE Access*, vol. 6, pp. 4207–4219, 2018.
[11] M. Zhang, H. Wang, and X. Ni
The LPI radar signal recognition based on deep learning and support vector machine
Sci. Technol. Rev., vol. 37, no. 4, pp. 69–75, 2019.
[12] D. Guo, B. Chen, W. Chen, C. Wang, H. Liu, and M. Zhou
Variational temporal deep generative model for radar HRRP target recognition
IEEE Trans. Signal Process., vol. 68, no. 1, pp. 5795–5809, 2020.
[13] L. Du, H. Liu, Z. Bao, and M. Xing
Radar HRRP target recognition based on higher order spectra
IEEE Trans. Signal Process., vol. 53, no. 7, pp. 2359–2368, Jul. 2005.
[14] L. Shi, P. Wang, H. Liu, L. Xu, and Z. Bao
Radar HRRP statistical recognition with local factor analysis by automatic Bayesian Ying–Yang harmony learning
IEEE Trans. Signal Process., vol. 59, no. 2, pp. 610–617, Feb. 2011.
[15] W. Kang, Y. Zhang, and X. Dong
Micro-Doppler effect removal for ISAR imaging based on bivariate variational mode decomposition
IET Radar, Sonar, Navigation, vol. 12, no. 1, pp. 74–81, 2018.
[16] L. Sun, X. Lu, and W. Chen
Joint sparsity-based ISAR imaging for micromotion targets
IEEE Geosci. Remote Sens. Lett., vol. 13, no. 11, pp. 1734–1738, Nov. 2016.
[17] C.-Y. Chong and S. P. Kumar
Sensor networks: Evolution, opportunities, and challenges
Proc. IEEE, vol. 91, no. 8, pp. 1247–1256, Aug. 2003.
[18] M. Xiang and C. Han
Global optimization for distributed and quantized Bayesian detection system
In Proc. Int. Conf. Inf. Fusion, 2000, pp. 658–663.
[19] Y. Wang and X. R. Li
Distributed estimation fusion with unavailable cross-correlation
IEEE Trans. Aerosp. Electron. Syst., vol. 48, no. 1, pp. 259–278, Jan. 2012.
[20] J. Wang, W. Yi, L. Kong, and S. Guo
Multi-frame detection method for distributed radar network
In Proc. IEEE Radar Conf., 2018, pp. 880–884.
[21] T. N. Kipf and M. Welling
Semi-supervised classification with graph convolutional networks
In Proc. Int. Conf. Learn. Representation, 2017, pp. 1–14.
[22] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini
The graph neural network model
IEEE Trans. Neural Netw., vol. 20, no. 1, pp. 61–80, Jan. 2009.
[23] S. Deng, S. Wang, H. Rangwala, L. Wang, and Y. Ning
Cola-GNN: Cross-location attention based graph neural networks for long-term ILI prediction
In Proc. Int. Conf. Inf. Knowl. Manage., 2020, pp. 245–254.
[24] A. Kapoor et al.
Examining COVID-19 forecasting using spatio-temporal graph neural networks
2020, arXiv:2007.03113.
[25] B. Yu, H. Yin, and Z. Zhu
Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting
In Proc. Int. Joint Conf. Artif. Intell., 2018, pp. 3634–3640.
[26] W. Chen, L. Chen, Y. Xie, W. Cao, Y. Gao, and X. Feng
Multi-range attentive bicomponent graph convolutional network for traffic forecasting
In Proc. AAAI Conf. Artif. Intell., 2020, pp. 3529–3536.
[27] S. Seo, A. Mohegh, G. Ban-Weiss, and Y. Liu
Automatically inferring data quality for spatiotemporal forecasting
In Proc. Int. Conf. Learn. Representation, 2018, pp. 1–11.
[28] S. Yan, Y. Xiong, and D. Lin
Spatial temporal graph convolutional networks for skeleton-based action recognition
In Proc. AAAI Conf. Artif. Intell., 2018, pp. 7444–7452.
[29] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun
Spectral networks and locally connected networks on graphs
In Proc. Int. Conf. Learn. Representation, 2014, pp. 1–14.
[30] 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.
[31] L. Yao, C. Mao, and Y. Luo
Graph convolutional networks for text classification
In Proc. AAAI Conf. Artif. Intell., 2019, pp. 7370–7377.
[32] D. Sui, Y. Chen, K. Liu, J. Zhao, and S. Liu
Leverage lexical knowledge for Chinese named entity recognition via collaborative graph network
In Proc. Conf. Empirical Methods Natural Lang. Process., 9th Int. Joint Conf. Natural Lang. Process., 2019, pp. 3830–3840.
[33] X. Liu, X. You, X. Zhang, J. Wu, and P. Lv
Tensor graph convolutional networks for text classification
In Proc. AAAI Conf. Artif. Intell., 2020, pp. 8409–8416.
[34] L. Liu, M. Ghogho, D. C. McLernon, and W. Hu
Pseudo maximum likelihood estimation of ballistic missile precession frequency
In Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2011, pp. 3796–3799.
[35] K. Cho et al.
Learning phrase representations using RNN encoder–decoder for statistical machine translation
In Proc. Conf. Empirical Methods Nat. Lang. Process., 2014, pp. 1724–1734.
[36] A. Hafizah, M. Hazwan, A. Emileen, R. Azmir, and S. Asem
RCS analysis on different targets and bistatic angles using LTE frequency
In Proc. Radar Symp., 2015, pp. 658–663.
[37] A. Swami and B. M. Sadler
Hierarchical digital modulation classification using cumulants
IEEE Trans. Commun., vol. 48, no. 3, pp. 416–429, Mar. 2000.
[38] E. Gökkaya and T. Günel
A novel hybrid approach for radar target classification based on SVM and central moments with PCA using RCS
In Proc. 11th Int. Conf. Elect. Electron. Eng., 2019, pp. 575–579.

Han Meng (Student Member, IEEE) received the B.S. degree in telecommunications engineering with management in 2018 from the Beijing University of Posts and Telecommunications, Beijing, China, where she is currently working toward the Ph.D. degree in information and communication engineering with the Key Laboratory of Universal Wireless Communications, Ministry of Education.
Her research interests include graph neural network and automatic aircraft recognition.

Yuexing Peng (Member, IEEE) received the Ph.D. degree in information and communication systems from Southeast University, Nanjing, China, in 2004.
From 2006 to 2008, he was a Postdoctoral Research Associate with the Beijing University of Posts and Telecommunications, where he has been a Faculty Member since 2008, and is currently an Associate Professor. He has authored or coauthored four books and more than 100 peer-reviewed papers in leading international journals and conferences, and holds 15 patents granted. His research interests include intelligent signal processing, machine learning, industrial Big Data, and Internet of Things.
Wenbo Wang (Member, IEEE) received the B.S. degree in communication engineering and the M.S. and Ph.D. degrees in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), in 1986, 1989, and 1992, respectively. From 1992 to 1993, he was a Researcher with ICON Communication Inc., Dallas, TX. Since 2000, he has been a Chair Professor with BUPT. Since 2002, he has been the Director of the Beijing Institute of Communication, and an Assistant Director of the National Defense Communication Committee, China Institute of Communication. His research interests include wireless communications, signal processing, and software defined radio.

Peng Cheng (Senior Member, IEEE) received the B.S. and M.S. degrees, with honors, in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2006 and 2009, respectively, and the Ph.D. degree in telecommunications from Shanghai Jiao Tong University, Shanghai, China, in 2013. From 2014 to 2017, he was a Postdoctoral Research Scientist with CSIRO, Sydney, Australia. From 2017 to 2020, he was an ARC DECRA Fellow/Lecturer with The University of Sydney, Australia. He is currently a Senior Lecturer (Tenured Associate Professor U.S. Systems) with the Department of Computer Science and Information Technology, La Trobe University, Bundoora, VIC, Australia, and is affiliated with The University of Sydney, Australia. He has authored or coauthored more than 70 peer-reviewed research papers in leading international journals and conferences. His research interests include wireless AI, machine learning, IoT, millimeter-wave communications, and compressive sensing theory. Dr. Cheng is an ARC DECRA Fellow.

Yonghui Li (Fellow, IEEE) received the Ph.D. degree in communications engineering from Beihang University, Beijing, China, in 2002. He is currently a Professor and the Director of Wireless Engineering Laboratory, The University of Sydney, Camperdown, NSW, Australia. He holds a number of patents granted and pending. His research focuses on wireless communications, including MIMO, millimeter wave communications, machine-to-machine communications, coding techniques, and cooperative communications. Prof. Li was the recipient of the Australian Queen Elizabeth II Fellowship in 2008, the Australian Future Fellowship in 2012, and several best paper awards from the IEEE conferences, such as the ICC and the PIRMC. He is an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was the Guest Editor for several IEEE journals.

Wei Xiang (Senior Member, IEEE) is currently the Cisco Research Chair of AI and IoT and the Director of the Cisco-La Trobe Centre for AI and IoT, La Trobe University, Bundoora, VIC, Australia. He was the Foundation Chair and the Head of the Discipline of IoT Engineering, James Cook University, Cairns, Australia. He has authored or coauthored more than 250 peer-reviewed papers, including three books and 200 journal articles. His research interests include the Internet of Things, wireless communications, machine learning for IoT data analytics, and computer vision. Mr. Xiang received the TNQ Innovation Award in 2016, the Pearcey Entrepreneurship Award in 2017, and the Engineers Australia Cairns Engineer of the Year in 2017. He was the co-recipient of four best paper awards at the WiSATS’2019, the WCSP’2015, the IEEE WCNC’2011, and the ICWMC’2009. He was awarded several prestigious fellowship titles. He was named a Queensland International Fellow (2010–2011) by the Queensland Government of Australia, an Endeavour Research Fellow (2012–2013) by the Commonwealth Government of Australia, a Smart Futures Fellow (2012–2015) by the Queensland Government of Australia, and a JSPS Invitational Fellow jointly by the Australian Academy of Science and Japanese Society for Promotion of Science (2014–2015). He was the Vice Chair of the IEEE Northern Australia Section from 2016 to 2020. He was the Editor for the IEEE Communications Letters (2015–2017), and is currently an Associate Editor for the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, the IEEE INTERNET OF THINGS JOURNAL, the IEEE ACCESS, and the Scientific Reports. He was a large number of international conferences in the capacity of General Co-Chair, TPC Co-Chair, Symposium Chair, etc. Due to his instrumental leadership in establishing Australia’s first accredited Internet of Things Engineering degree program, he was inducted into Pearcy Foundation’s Hall of Fame in October 2018. He is a TEDx Speaker and an Elected Fellow of the IET in U.K. and Engineers Australia.