A General Recognition Approach for Formation Change of Group Targets

Dan Wang(✉), Yongfu Wang, Ke Li, Yufei Huang, and Luyuan Wang

China Academy of Space Technology (CAST), Beijing Institute of Spacecraft System Engineering (ISSE), Beijing, China
wangdan_ict_hit@163.com

Abstract. The carrier strike groups (CSGs) are the major components of the national navy. The surveillance of CSG formation changes and other behaviors is critical to identify the other party's behavioral intentions and win the first opportunity in the battle field. As the spread range of a CSG is large, satellite remote sensing imagery is the most effective way to capture the CSG behavior change. However, due to the difference of satellite imaging angle of view, the apparent characteristics of similar formation may vary greatly, which may lead to the decline of formation change recognition performance. Inspired by the fact that a CSG formation is usually composed of multiple basic units, which appears a strong spatio-temporal relationship between each other, this paper proposes a graph model based method to identify the formation changes of group targets. Firstly, the graph model method is proposed to model the spatial structure relationship of the group targets. Secondly, multi-scale convolution kernel features and spatio-temporal graph convolution features are calculated to capture the formation sequential changes. Thirdly, a graph based, deep recursive convolution neural network model is learned and used to recognize formation changes. In our experiments, the formation transformation refers to the sequential transformation between two of seven specified formations. Experimental results show that the graph model and the spatio-temporal graph convolution feature representation are efficient and robust, to describe formation trajectory of the simulated data. Meanwhile, the recursive convolution neural network model greatly improves the recognition performance of target formation change, which can provide necessary technical support for correct decision-making and occupying the initiative in the battlefield.

Keywords: Group Targets · Graph Model · Deep Learning

1 Introduction

Group behavior analysis and recognition has attracted more and more attention from researchers [1–4]. With the continuous maturity of computer vision technology, the research on this topic has been deepened [5–7]. Group targets can be divided into two categories: one is dense target groups, such as crowds gathered in various occasions such as shopping malls; the other is sparse target group, such as athletes on the stadium,

H. P. Urbach and H. Jiang (Eds.): ISSOIA 2022, SPPHY 295, pp. 129–137, 2023.
https://doi.org/10.1007/978-981-99-4098-1_13

carrier battle group on the sea, aircraft group in the air, satellite group in space, etc. This paper focuses on the behavior analysis of sparse target groups, especially the formation change of groups. Taking the carrier strike group as an example, it is a typical sparse target group, which can be seen as composed of some discrete points. The change of array formation often indicates the change of a certain battlefield situation, and is the basis for identifying behavior intention of the enemy and winning the battlefield first opportunity [8, 9]. Consequently, it is of great application value to study the identification method of array formation changes.

2 Related Works

The CSGs are the main organization of national navy. By monitoring the CSG with satellites, identifying its formation and change behavior, it can provide decision-making support to military commanders, and even determine the outcome of a campaign. Due to the wide spread range of CSGs, the most effective way to obtain the navigation of the opposite CSGs in the actual battlefield is satellite imagery. However, the research work on predicting formation changes based on satellite detection results is relatively limited.

Deng et al. [9] proposed MVC (Multi view Point Context) descriptor based on Archimedes spiral for ship formation recognition, which introduced a probability density function of observation points, and proposed a similarity measurement method to directly identify ship formation. However, the descriptors are lack of spatial and temporal structure information. The paper [10] regarded the formation as a scene in two-dimensional space, in which the known formation is assumed as a template, and the tested ones can be recognized by calculating the scene similarity with the template. The method requires target locations and less target positioning accuracy, but has poor deformation resistance. The works in [1] and [11] mainly focus on the design of linear formation. In [11], an algorithm for recognition of warship formation has been studied based on Hough transform technology. When the target information is seriously polluted, the improved K-means clustering algorithm is further used to cluster the local peaks of the accumulation matrix obtained by Hough transform. Finally, the parameters of the formation are accurately extracted, according to the peak clustering results. However, limited by the design of line detection, its adaptability is not strong. In addition, some researchers use graph models composed of a series of key points to identify object types, and even array formation.

Generally, there are several basic types of fleet formation, based on unit structures of fleet groups. By analyzing the basic units in the formation and inferring the formation composition, the goal of identifying the formation can be achieved. It is difficult to find a unified approach to recognize group targets, when they are rotated, scaled and deformed. Thus some scholars have utilized deformation ability of graph models to build target recognition algorithms. In this paper, a graph based deep network learning method is proposed. By combining the graph model and deep convolution neural networks, we proposed an efficient and robust graph feature representation method for feature extraction and analysis of large-scale formation trajectory data. Robust dynamic graph prediction method is also used to achieve dynamic real-time prediction and analysis of formation trajectory data.

In this paper, the typical sparse group of target carrier strike group is taken as an example to study two aspects: (1) Extraction both of static features and dynamic spatio-temporal changes of the carrier strike group formation; (2) Identification and analysis of aircraft carrier strike group formation change. Since the ship formation is usually composed of multiple basic units, there is a strong spatial relationship between ships. Therefore, a graph model based on spatiotemporal trajectory data is constructed to describe the spatio-temporal correlation and dependency between the formation change data, which has a strong anti deformation ability. By establishing the recursive convolution neural network (RCNN) model of dynamic graph, the problem of joint modeling of temporal motion and spatial structure is solved, and the recognition accuracy of group target formation changes is improved.

3 Proposed Method

3.1 Graph Model Construction of Group Targets

Given a remote sensing image, a triplet can be constructed according to the ship group attributes in the image, which represents the graph model $\mathcal{G} = (\mathcal{V}, \mathbf{X}, \mathbf{A})$ where $\mathcal{V}$ represents the set of nodes, $\mathbf{X}$ represents the node features and $\mathbf{A}$ represents the adjacency matrix. In the node set $\mathcal{V} = \{v_1, \ldots, v_n\}(|\mathcal{V}| = n)$, we assume each ship as a node and extract features according to the node attributes, including the ship position, speed, dwell time etc. Formally, $\mathrm{x}_i = f(x, v_i) \in \mathbb{R}^{n \times d}$, where x_i represents the i-th node, f represents the feature extractor, and d represents the data dimension. By concatenating node features, we can get the feature matrix composed of all node features: $\mathrm{X} = [\mathrm{x}_1^{\mathrm{T}}; \mathrm{x}_2^{\mathrm{T}}; \ldots; \mathrm{x}_{\mathrm{n}}^{\mathrm{T}}] \in \mathbb{R}^{\mathrm{n} \times \mathrm{d}}$. For simplicity, $\mathbf{X_i}$ and x_i will not be distinguished later, which represent the i-th row in characteristic matrix or i-th node features, with the same meaning.

Assume the edge set ε consists of the edges connecting all nodes in graph $\mathcal{G}$. The edge construction depends on the spatial, semantic or other potential influence factors between nodes. We formally define the edge between v_i and v_j node as $e_{ij} = \preccurlyeq_\epsilon(\mathcal{K}(v_i, v_j))$, where $\preccurlyeq_\epsilon(\cdot)$ is the indicator function with the threshold ϵ, and $\mathcal{K}$ is the similarity measurements, such as Euclidean distance. For example, the edge value is 1, if the distance $\mathcal{K}$ between v_i and v_j is greater than the threshold value ϵ, otherwise it is 0. All connection node pairs form an edge set $\varepsilon = \{(v_i, v_j)|e_{ij} = 1, v_i, v_j \in \varepsilon\}(|\varepsilon| = m)$, which can also be represented as an unweighted adjacency matrix $\mathbf{A} \in \mathbb{R}^{\mathrm{n} \times \mathrm{n}}$,

$$A = (e_{ij})_{n \times n}, e_{ij} = \begin{cases} 1, (v_i, v_j) \in E; \\ 0, \textit{otherwise}. \end{cases}$$

In addition, for the relationship with edges connected, we use features to enhance edge information. The most common method is to introduce weighted edges. The higher the correlation between two nodes, the greater the corresponding edge weight. We can use measurement function $\mathcal{K}$ to define $\mathrm{A_{ij}} = e_{ij} \times \mathcal{K}(v_i, v_j)$. For example, only the adjacent edges have non-zero weights, so the adjacency matrix $\mathbf{A}$ not only records the adjacency of nodes, but also describes the correlation degree of adjacent nodes.

Based on adjacency matrix **A**, Laplacian matrix can be constructed as a matrix representation of node relationships in the graph. Given a graph model $\mathcal{G} = (\mathcal{V}, \mathbf{X}, \mathbf{A})$, its Laplace matrix can be defined as

$$L = D - \mathbf{A},$$

where D is the degree matrix of the graph, expressed as a diagonal matrix $D = diag[d_1, d_2, \ldots, d_n]$ and $d_i = \sum_j A_{ij}$.

3.2 Spatial-Temporal Graph Convolution (STGC) Model Based Group Target Formation Change Recognition

In this paper, the ship formation is modeled as an undirected attribute graph $\mathcal{G} = (\mathcal{V}, \mathrm{A}, \mathrm{X})$, with a set of N vertices $\mathcal{V} = \{v_n\}_{n=1}^{N}$, a weighted adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ is, and a signal matrix on the vertices $\mathbf{X} \in \mathbb{R}^{N \times d}$. So the formation graph sequence with the length of T can be represented as $(\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_T)$, where $\mathcal{G}_k = (\mathcal{V}_k, A_k, X_k)$ represents the formation on the k-th slice. Next, we will introduce the Spatial Temporal Graph Convolution (STGC) model for dynamic graph sequence analysis.

3.2.1 Multiscale Graph Convolution Kernel

Since the image has a grid structure, the receptive field of standard convolutional neural network (CNN) on the image can be easily defined as a local rectangular region. So convolution filtering on regular structured data is easy to operate. On the contrary, it is relatively difficult to construct convolution kernels on irregular structured graphs, mainly because homogeneous graph structures correspond to the same filtering response. Inspired by graph theory, with the help of the adjacency matrix **A**, which represents the connection relationship between the vertices of a graph. A_k accurately records the vertices that can be reached by k step by step connection on the graph. Therefore, the receptive field of k neighborhood $\psi_k(\mathbf{A})$ can be constructed by defining k order polynomial of **A**. The simplest strategy is adopted here, $\psi_k(\mathbf{A}) = \mathbf{A}_{\boldsymbol{k}}$. In practice, the Laplacian matrix L of the graph can be used instead of **A** to avoid being affected by the scale of matrix norm in recursive reasoning. Therefore, for the local receptive field with a scale of K, the following multi-scale convolution filtering is defined:

$$Z = \mathcal{G} * f = \sum\nolimits_{k=0}^{K-1} \psi_k(L) X V_k,$$

where $\psi_k(L)$ represents the receptive field of k scale, $V_k \in \mathbb{R}^{d \times d'}$ is the corresponding signal transformation matrix. The vertex information in all the k scale receptive fields is weighted and synthesized by calculating $\psi_k(L)X$. Consequently, for the graph, its filtering response is homogeneous and invariant.

3.2.2 Spatio-Temporal Graph Convolution

Base on the above multi-scale graph convolution kernel, and further inspired by the design concept of Auto Regressive Moving Average (ARMA) model, the following

spatio-temporal graph convolution model is built:

$$Y_{t+1} = \sum_{k=0}^{K_1-1} \psi_k(L) Y_t W_k + X_t V_0,$$
$$O_{t+1} = Y_{t+1} + \sum_{k=1}^{K_2-1} \psi_k(L) X_t V_k,$$

where, $\psi_k(\cdot)$ is the receptive field function of k-th scale, $\{\mathbf{W}_k \in \mathbb{R}^{d' \times d'}, \mathbf{V}_k \in \mathbb{R}^{d \times d'}\}$ is the signal transformation matrix of the k-th scale,. In the above model, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{d'}]$ and $\mathbf{O} = [\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_{d'}]$ can be regarded as hidden state and output state respectively. Along the time dimension, the signal Y_{t+1} is regressed recursively by the local convolution kernel in the above equation, so that the dynamic changes can be serialized and encoded.

The output signal O_{t+1} in the above equation integrates the spatial graph convolution signal and the dynamic sequential signal, and each output signal $\mathbf{o}_i$ depends on all input signals $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d\}$. In particular, when signals are independent and the spatial-temporal convolutional filter operates on the signal channel respectively, the dynamic graph convolutional model can be written as:

$$Y_{t+1} = \sum_{k=0}^{K_1-1} \psi_k(L)\, Y_t diag(w_k) + X_t diag(v_0),$$
$$O_{t+1} = Y_{t+1} + \sum_{k=1}^{K_2-1} \psi_k(L)\, X_t diag(v_k),$$

where $\mathbf{w}_k = [w_{k1}, w_{k2}, \ldots, w_{kd}]^{\mathrm{T}}$ and $\mathbf{v}_k = [v_{k1}, v_{k2}, \ldots, v_{kd}]^{\mathrm{T}}$ are mapping parameters, and w_{ki} and v_{ki} is related to k-th scale of the i-th signal. Note that it is assumed that the number of output dimensions is the same as the one of the input. In the case of signal independence, it is easy to expand to $d' \neq d$.

3.2.3 Construction of Deep Graph Convolution Neural Network Model

The above recursive convolutional model can be easily extended to the deep structures. Specifically, the recursive model is used as a basic layer to stack into a multi-layer network architecture, where the output signal $\mathbf{O}$ of the bottom layer is used as the input of the top layer. Formally,

$$Y_{t+1}^{(l)} = \sum_{k=0}^{K_1-1} \psi_k(L) Y_t^{(l)} W_k^{(l)} + O_t^{(l-1)} V_0^{(l)},$$
$$O_{t+1}^{(l)} = Y_{t+1}^{(l)} + \sum_{k=1}^{K_2-1} \psi_k(L)\, O_t^{(l-1)} V_k^{(l)},$$

where l represents the number of network layers. With the increasing number of layers, the receptive field scale of the convolution kernel can become larger, so the top layer can extract more global information. In other words, the recursive model is regarded as a basic neural network layer, and a multi-layer network structure is formed by stacking more layers. The output signal of the current layer can be used as the input signal of the next layer. As the number of network layers increases, the receptive field of convolution kernels becomes larger, so that the top layer of the network can capture global information. The framework is shown in the figure below (Fig. 1).

Different basic features in the graph are extracted by constructing multi-channel local convolution kernels, while higher level features are extracted by designing multiple

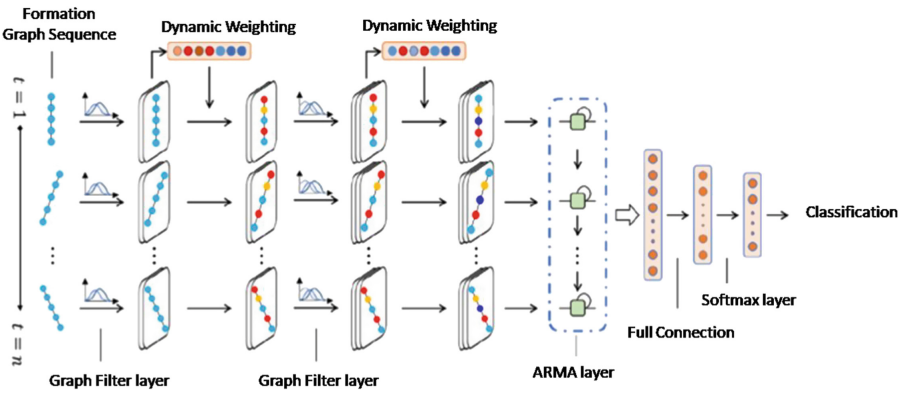

Fig. 1. Spatio-Temporal Dynamic Graph Network Framwork

convolutional layers. Gradient descent method is adopted to learn the model parameters. By using k-neighborhood local filters, the learning complexity is independent of the number of input samples and only depends on $\boldsymbol{O}(k)$. We consider of the maximum pooling operation, local average pooling, or constructing a binary tree according to some rules, to merge some similar vertices for down sampling. Finally, with a full connection layer, the learned feature representation is mapped into the sample label space, to classify the traveling trajectory.

4 Experiments

4.1 Dataset

The formation change in this paper refers to the sequence transformation between two of seven formations, thus there are 42 transformation types of sequences. Each sequence has 100 frames; the pixel difference between adjacent frames is the average difference of the formation distance between the two ends of the sequence. During the experiments, 4200 sequences of training set and 4200 sequences of test set are simulately generated, that is, 100 sequences of each formation change in training set and test set respectively. Among them, several formation transformation diagrams are shown in Fig. 2.

4.2 Comparisons

The proposed STGC model is experimentally compared with the latest methods. The experimental results are shown in Table 1. Among them, P-LSTM is a model for emotion classification based on short-term memory recurrent neural network. Comparatively, STGC has greatly improved the recognition accuracy (5.5%). Cov-RP proposed a new region feature descriptor, the covariance matrix feature, to represent the region of interest, which is applied to target detection and texture classification. Ker-RP proposes an open framework, which uses the kernel matrix on feature dimension as a universal representation, and explores a more appropriate representation based on symmetric positive

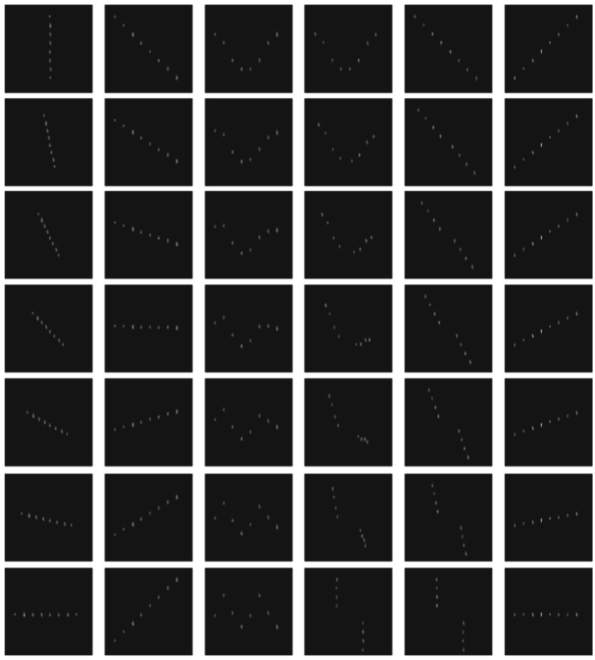

Fig. 2. Example of formation change. From left to right, transforms are shown in columns. (a) from columns to rows, (b) from right echelon to left echelon, (c) from V-shaped echelon to wedge-shaped echelon, (d) from V-shaped echelon to interlaced echelon, (e) from right echelon to interlaced echelon, (f) from left echelon to rows.

definite matrix, to obtain better recognition performance. RSR-ML can deal with high dimensional symmetric positive definite matrices, by constructing a low dimensional and more discriminative symmetric positive definite manifold. Thus it adopts standard orthogonal projection to model the mapping from high-dimensional symmetric positive definite manifold to low dimensional positive definite manifold, to improve the recognition accuracy.

Table 1. Comparison results of STGC model

Methods	Recognition rates
P-LSTM	92.6%
Cov-RP	92.4%
Ker-RP	95.2%
RSR-ML	96.8%
STGC	**98.1%**

Based on the above experiments, we extract 75%, 50% and 25% sequence fragments from each complete sequence as a new test set and use the complete sequence fragments in the training set. Comparisons are shown below, which realizes the sequence type prediction, according to incomplete sequence frames.

Table 2. Recognition rate of observed sequence data with different proportion

Proportion of observed sequences	Recognition rate
25%	87.5%
50%	92.4%
75%	97.8%
100%	98.1%

Table 2 shows, it will bring a great reduction 9.6% on the accuracy of the final prediction, when using only the first 25% of the full sequence, while it will bring less impact on the results, when using more than 75% of the full sequence. It demonstrates the robustness of the proposed model.

5 Conclusion

This paper proposes an effective graph structured modeling method for the formation change of group targets, based on the deep graph network model. By comprehensively considers the spatio-temporal data information of the dynamic change of group targets, we focus on the problem of efficient and robust graph feature representation. A spatio-temporal graph convolution method is proposed to extract features from large-scale formation trajectory data, which greatly improves the discrimination ability and accuracy of the deep graph network. Finally the dynamic identification and analysis of formation change data of ship group targets and experimental results demonstrate the effectiveness of the proposed method.

References

1. Bing, Q., Bin, X.: Research on automatic recognition of warship formation based on template matching. Comput. Simul. **23**(9), 4–6 (2006)
2. Lan, T., Wang, Y., Yang, W., Ro-binovitch, S.N., Mori, G.: Discriminative latent models for recognizing contextual group activities. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 1549–1562 (2012)
3. Zhang, Y., Dong, S., Bi, K.: Recognition algorithm of ship formation based on Hough transform and clustering. J. Mil. Ind. **37**(4), 648–655 (2016)
4. Deng, Z., Vahdat, A., Hu, H., Mori, G.: Structure inference machines: recurrent neural networks for analyzing relations in group activity recognition. In: IEEE International Conference on Computer Vision and Pattern Recognition (2016)

5. Ibrahim, M.S., Mori, G.: Hierarchical relational networks for group activity recognition and retrieval. In: European Conference on Computer Vision (2018)
6. Wu, J., Wang, L., Wang, L., Guo, J., Wu, G.: Learning actor relation graphs for group activity recognition. In: IEEE International Conference on Computer Vision and Pattern Recognition (2019)
7. Azar, S.M., Atigh, M.G., Nickabadi, A., Alahi, A.: Convolutional relational machine for group activity recognition. In: IEEE International Conference on Computer Vision and Pattern Recognition (2019)
8. Xiaochun, Z.: Analysis of US aircraft carrier formation and anti submarine capability. Ship Sci. Technol. **35**(9), 143–148 (2013)
9. Deng, C., Cao, Z., Xiao, Y., Chen, Y., Fang, Z., Yan, R.: Recognizing the formations of CVBG based on multiviewpoint context. IEEE Trans. Aerosp. Electron. Syst. **51**(3), 1793–1810 (2015)
10. Shouquan, D., Wei, S.: Formation recognition algorithm based on spatial direction similarity. Fire Command Control **35**(11), 167–169 (2010)
11. Leng, H., Guan, Q., Wu, X.: Recognition of maritime formation linetype based on domain knowledge. Ship Sci. Technol. **35**(2), 103–106 (2013)