frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 24 March 2025
DOI 10.3389/fphy.2025.1538605

Check for updates

OPEN ACCESS

EDITED BY
Gaogao Dong,
Jiangsu University, China

REVIEWED BY
Lucas Bondan,
National Education and Research Network, Brazil
Abdelkarim Ben Sada,
University College Cork, Ireland
Maythem Derweesh,
Mustansiriyah University, Iraq

*CORRESPONDENCE
Yucai Zheng,
s1327699@live.hkmu.edu.hk

RECEIVED 03 December 2024
ACCEPTED 26 February 2025
PUBLISHED 24 March 2025

CITATION
You Z and Zheng Y (2025) Security anomaly detection for enterprise management network based on attention mechanism. *Front. Phys.* 13:1538605. doi: 10.3389/fphy.2025.1538605

COPYRIGHT
© 2025 You and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Security anomaly detection for enterprise management network based on attention mechanism

Zhaohan You[1] and Yucai Zheng[2]*

[1]Hong Kong Baptist University, College of Communication, Interactive Media Specialty, Kowloon, Hong Kong SAR, China, [2]Hong Kong Metropolitan University, School of Science and Technology, Hong Kong SAR, China

With the rapid growth of data volume in enterprise management systems and the continuous complexity of network architecture, traditional network security protection methods are no longer sufficient to fully address the security challenges. In response to the problems of insufficient accuracy and high time consumption in traditional network security anomaly detection methods, this paper proposes a detection model combining attention mechanism based spatial convolutional network and gated attention transformer (AMSCN-GADetector). It is an enterprise management network security anomaly detection method based on deep learning, aiming to achieve efficient and intelligent monitoring and management of security anomaly data in enterprise management network. This method combines spatial convolutional network and gating mechanism, which are used to extract spatial features from enterprise management network security data and learn non-local interaction relationships between features. In addition, by introducing attention mechanism, AMSCN-GADetector can accurately calculate the importance weights of network security data features. This effectively reduces the loss of critical security information in the detection process. Finally, through comparative experiments, it is verified that AMSCN-GADetector exhibits superior detection performance compared to other models, providing solid technical support for the stable operation and long-term development of enterprise management.

KEYWORDS

enterprise management network, attention mechanism, security, deep learning, anomaly detection

Highlights

(1) This paper proposes a detection model for enterprise management network security anomaly detection.
(2) The method integrates gating mechanism with self-attention mechanism, which can accurately model contextual semantic information and global spatial relationships.
(3) The method demonstrates advantages in detection accuracy and attack threat type recognition ability.

1 Introduction

With the rapid advancement of information technology, enterprise management systems have become the central hub for intelligent decision-making in modern business operations. They not only gather massive amounts of business data, but also leverage cutting-edge information technology to achieve real-temporal data access and efficient governance. However, in this wave of digital transformation, enterprise management systems are also facing increasingly complex and ever-changing network security challenges. Threats such as system vulnerabilities, hacker infiltration, malware attacks and sensitive data leaks are like hidden currents surging, constantly testing the bottom line of stable operation and data security for enterprise [1].

In the early stages of enterprise management network architecture, people often focus on the ease of use and accessibility of the network, while the emphasis on network security appears relatively lagging behind. When enterprise networks are still limited to a small circle of internal LAN, network security issues have not yet surfaced. But with the acceleration of enterprise digital transformation, comprehensive network information services have gradually become standard. The security of data in the network, whether stored statically on servers or dynamically transmitted within the network, constitutes the core of network security. Network security is a field that integrates the wisdom of multiple disciplines such as computer science, network technology, communication technology, information security technology, cryptography, applied mathematics, number theory and information theory [2–4]. Its core mission is to safeguard the integrity and security of network data.

The enterprise management network system, with its wide distribution, open architecture, resource sharing and common channels, greatly enhances the practicality of the network. But it also invisibly increases the vulnerability of the system, forcing enterprises to face the serious issue of network security. The rampant spread of network viruses, cunning attacks by hackers and employees' deep expectations for network security jointly promote the enterprise's network security protection to a higher level [5–7]. Enterprise network security aims to ensure that all components of the network system are protected from accidental or malicious attacks, while ensuring the continuous and stable operation of the system. Its ultimate goal is to achieve high confidentiality, integrity, availability and controllability in the process of information processing and transmission.

The enterprise management system, as a key support for the operation of enterprises in the new era, is also the focus of global business competition and technological innovation. However, in the early stages of enterprise management system networking and informatization construction, people focus more on its availability and convenience, while neglecting the security of network information resources. With the continuous iteration and upgrading of enterprise management systems, enterprises begin to provide comprehensive network information services and resource sharing. However, the openness of system structure, wide distribution, common channels and resource sharing force enterprises to face more severe network security challenges. Security incidents such as data loss, information leakage, system tampering, software vulnerabilities, hacker attacks, malicious code and internal personnel misoperation are like temporal bombs, seriously threatening the normal operation of enterprises and causing huge losses to enterprises and users [8–10]. Therefore, conducting in-depth research on network factors that pose a threat to enterprise management systems and building a comprehensive security system has become a major issue related to the construction and development of enterprise systems.

Deep learning, as a key technology in the field of artificial intelligence, is highly prominent in areas such as image recognition, speech recognition and natural language processing due to its powerful data processing and self-learning capabilities [11]. In the field of network security, deep learning technology has also demonstrated extraordinary strength and is widely used in scenarios such as network attack detection, malware identification and anomaly traffic analysis. Especially the introduction of attention mechanism makes it more precise and efficient in network security detection. The attention mechanism can simulate the attention allocation in the model, focusing on key information and ignoring irrelevant details, thereby significantly improving the accuracy and efficiency of detection. Through the comprehensive application of anomaly detection, intrusion prevention and risk assessment technologies, combined with attention mechanism deep learning technology, it is possible to create a network security defense line for enterprise management systems, ensuring their stable operation and data security.

To effectively respond to network security attacks and intrusions, the network security detection system, as an intelligent protection system that combines firewall defense and real-temporal attack judgment and defense functions. It not only strengthens the internal protection capabilities of the network, but also enhances the level of resistance to external attacks. Especially the network security detection system based on attention mechanism can more accurately identify and resist network attacks, providing a more solid guarantee for network security. In response to the performance limitations, high false alarm rates and long detection cycles of traditional machine learning methods in network security detection, more and more researchers are introducing deep learning into the field of network security data detection and using attention mechanisms to further improve detection performance [12]. In deep learning, the addition of attention mechanisms enables it to more accurately capture key information and improve detection accuracy. By constructing different neural network structures and optimizing feature extraction, combined with attention mechanisms, better classification prediction models can be obtained, providing stronger support for network security protection [13].

Although traditional deep learning methods are applied in network security data detection, these methods still seem inadequate when faced with professional enterprise management of network security data. Therefore, applying deep learning, especially combining attention mechanisms, to the field of enterprise management network security data detection to achieve more accurate identification and management has become an effective and promising solution. Deep learning can extract nonlinear features from massive and complex data, forming higher-level data representations. While attention mechanisms can focus on key information and ignore irrelevant details, thereby further improving learning effectiveness. Therefore, the application of attention mechanism in the field of network security data detection

to manage enterprise management systems has become an effective method. Our main contributions are summarized as follows.

(1) In response to the challenges of low efficiency, low detection accuracy and inability to accurately identify attack threat types in current network security data detection methods based on deep learning, this paper proposes AMSCN-GADetector for enterprise management network security anomaly detection. The preliminary feature extraction of enterprise network security data is carried out using spatial convolutional network, which effectively extracts the spatial structure and local features of the data. Subsequently, the features output by the convolutional layer are used as inputs for the attention mechanism to dynamically adjust the importance of features, enabling the model to focus on key information.
(2) Furthermore, AMSCN-GADetector integrates gating mechanism with self-attention mechanism, which can accurately model contextual semantic information and global spatial relationships. This ensures consistency and accuracy of feature representation, effectively reduces the lack of key information in enterprise management network security data and significantly improves the accuracy of model detection.
(3) In order to verify the performance of AMSCN-GADetector, experimental comparisons are conducted on multiple network security datasets. The results show that compared with other baseline methods, AMSCN-GADetector demonstrates advantages in detection accuracy and attack threat type recognition ability, providing strong support for enterprise management network security protection.

The rest of this article consists of four parts. Section II reviews related literature. Section III provides a detailed introduction to the enterprise management network security anomaly detection method based on AMSCN-GADetector. Section IV analyzes the comparative effect of AMSCN-GADetector on network security anomaly detection through experiments and metrics based on multiple datasets and baselines. Finally, Section V is the summary.

2 Literature review

In the management of enterprise network environments, due to the large amount and high complexity of data, traditional machine learning methods were inadequate for network security anomaly detection. Therefore, deep learning technology had gradually become a research hotspot in this field. Du et al. [14] proposed a deep long short-term memory network model called DeepLog, whose core lied in the long short term memory (LSTM). This could convert complex system logs into easily processed natural language sequences. Its advantage lay in its ability to automatically learn and capture inherent pattern features from normal running log data. Once the system logs deviated from these normal patterns, DeepLog could effectively identify and mark them as anomaly, achieving early warning. In addition, DeepLog had the ability for online incremental learning, which could automatically adjust and optimize model parameters as new log data continues to flood in. This ensured that the model could keep up with system changes and identify new log patterns in a temporally manner. Javaid et al. [15] proposed an intrusion detection method based on sparse autoencoder and softmax regression. The method first used sparse autoencoder to extract features in an unsupervised manner, and then used softmax regression algorithm to construct a classifier to detect anomaly traffic in the network. The experimental results on NSL-KDD showed that the method performed well in accuracy, precision, recall and f-measures values. Shaikh et al. [16] proposed a deep learning framework that combined autoencoder (AE) and recurrent neural networks for access control of network traffic. This framework utilized AE to preprocess data and trained classification models using LSTM. Experimental results showed that this method performed well in reducing false positive rates. Su et al. [17] proposed a traffic anomaly detection model called BAT to address the issues of insufficient accuracy and complex feature engineering in the field of network security anomaly detection. This model integrated a bidirectional long short-term memory network (BiLSTM) with attention mechanism, aiming to optimize the extraction and utilization of network traffic features. The introduction of attention mechanism enabled the model to finely screen the data packet vector sequence processed by BiLSTM, focusing on key information. At the same time, BAT also integrated multi-layer convolutional neural networks (CNNs) to deeply explore local features of data and improve the accuracy of characterizing network traffic behavior. The application of softmax classifier achieved accurate identification of network traffic. As an end-to-end solution, BAT could automatically learn and abstract hierarchical key features from raw data, simplifying the model construction process and improving generalization ability. The experimental results on NSL-KDD showed that the accuracy of BAT reached 84.25%, demonstrating significant advantages in anomaly detection accuracy, efficiency and robustness.

Deep learning performed well in detecting network security anomaly in enterprise management. Derhab et al. [18] proposed an intrusion detection method based on temporal convolutional neural networks, which integrated temporal convolution into the convolutional neural network and optimized it for IoT data, improving accuracy and reducing training temporal. El Sayed et al. [19] proposed an improved deep neural network (DNN) and combined it with feature engineering to select the optimal subset for training, which improved the ability to detect anomaly traffic. Ma et al. [20] designed a feature extraction algorithm in the context of the Internet of Vehicles and constructed an intrusion detection system that could be deployed in vehicles using a lightweight gated recurrent unit (GRU). The experimental results showed that the method had high accuracy and real-temporal performance. Van et al. [21] used an improved deep learning model to extract features from TCP/IP traffic data for anomaly detection in industrial IoT data, experimentally verified the detection performance of AE and CNNs. Combining multiple different basic neural network structures can further enhance feature extraction performance. Andresini et al. [22] designed a multi class deep learning combination model based on convolutional layers and attention layers, utilizing attention mechanisms to enhance feature map extraction. The experimental results showed that the model made relatively accurate classification decisions for traffic data. Cao et al. [23] first used random forest and Pearson analysis for feature selection. Then they designed a combined network model combining TCN and GRU, embedded attention mechanism to assign weights to features. The high accuracy of

their method was experimentally verified. Lee et al. [24] proposed a model that combined AE and generative adversarial network (GAN) to address data imbalance and high-performance intrusion detection issues. Yao et al. [25] used the XGBoost algorithm to eliminate redundant features and combined CNN and Transformer to construct feature associations, improving accuracy. Elsayed et al. [26] proposed a hybrid model of BiLSTM and CNN for detecting anomaly in smart home networks, using CNN and BiLSTM to extract spatiotemporal features and achieving good detection results. Shone et al. [27] proposed a classification model for asymmetric deep AE and conducted experiments on KDDCup99 and NSL-KDD, demonstrating the potential for improving AE. Although these methods performed well in specific scenarios such as the IoT, the Internet of Vehicles and the Industrial Internet of Things, each method was designed for specific types of data or network environments. Therefore, their generality and generalization ability might be limited, making it difficult to directly apply them to other types of data or network environments.

3 Deep learning driven enterprise management network security anomaly detection method

3.1 AMSCN-GADetector

The use of deep learning to achieve intelligent detection of network security anomaly in enterprise management is an important research problem. At present, research on using deep learning for enterprise management network security anomaly detection is still in its preliminary stage. The existing deep learning-based methods for detecting network security anomaly in enterprise management have the problem of insufficient detection capability, with high false alarm and missed alarm rates, but low accuracy. Therefore, further research is needed for detecting network security anomaly in enterprise management based on deep learning. The use of spatial convolutional network and attention mechanism can better extract security anomaly features of enterprise management network and achieve intelligent detection of security anomaly in enterprise management network. Therefore, this paper proposes AMSCN-GADetector, it is based on attention mechanism, which combines spatial convolutional network with gated attention transformation units. This method incorporates an attention mechanism module that can dynamically reassign the importance weights of different network data features, enabling focus on information that is crucial for identifying security anomaly. In this way, AMSCN-GADetector not only reduces the burden of computing resources, but also significantly improves the accuracy and efficiency of enterprise management network anomaly detection.

To minimize the impact on neural network, we first need to perform data preprocessing, remove annotations and non ASCII characters, unify variable and function names. The purpose of this step is to improve the generalization ability of the model, as annotations, non ASCII characters, variable names and function names are not related to network vulnerabilities. Then we use the word2vec tool to encode it into vector form. As the neural network takes fixed length vectors as input, we need to do some data normalization processing on the encoded vectors. We mainly consider the following two situations, if the length of the transformed vector is greater than t, we delete the excess parts. Otherwise, we treat the missing values by filling the beginning or end of the vector with a zero vector.

In the field of enterprise management network security anomaly detection, the current development of deep learning technology shows a significant trend of continuously expanding dataset size and significantly increasing model complexity, which directly leads to a sharp expansion of model parameter quantity. Given the massive nature of enterprise network data, processing all traffic information comprehensively and indiscriminately will inevitably lead to a sharp increase in computing costs. Therefore, as an efficient information processing strategy, attention mechanism is particularly important in network security anomaly detection. This mechanism can intelligently guide the detection model to focus on key information segments in network data, thereby achieving precise configuration and efficient utilization of computing resources. This strategy not only significantly optimizes the execution efficiency of anomaly detection algorithm, but also further enhances the accuracy of network security anomaly recognition by effectively filtering redundant information.

The attention mechanism can be divided into two types based on its operating mode [28]. They are hard attention and soft attention. The hard attention mechanism adopts a strict screening strategy, selecting the most significant item from numerous network traffic information for individual processing, such as focusing on the traffic features with the highest probability for analysis. In contrast, soft attention mechanisms exhibit higher flexibility by assigning different weights to all traffic information, achieving reintegration and emphasis of information. Information with higher weights receives greater attention and weighted processing. We adopt the soft attention mechanism, with input and output dimensions consistent with CNN. This feature enables the module to be a plug and play component, flexibly and seamlessly embedded into existing CNN architecture, providing strong support for enterprise management network security anomaly detection tasks.

The basic network components of AMSCN GADetector are shown in Figure 1, which integrates a spatial convolutional network with a gated attention transformer unit, aiming to comprehensively and efficiently solve a series of weaknesses and limitations in security anomaly detection for enterprise management network. This method first extracts fine spatial structural features from the skeleton sequence of input enterprise management network security abnormal network traffic or log data through spatial convolution layers, which reflect the spatial distribution pattern of network activity. Subsequently, in order to further optimize the feature extraction process and improve the accuracy of detection, an attention mechanism module is introduced. It integrates spatial attention, temporal attention and channel attention mechanism to ensure that AMSCN-GADetector can focus on key network security event information in all aspects, thereby enhancing the model's understanding of complex network attack patterns or abnormal behavior features, promoting effective interaction and fusion between different features. Subsequently, the gated attention transformer unit is used to learn the non local interactions between these features, namely, the correlations between different network components or events, which further enhances the model's ability to identify network security abnormal

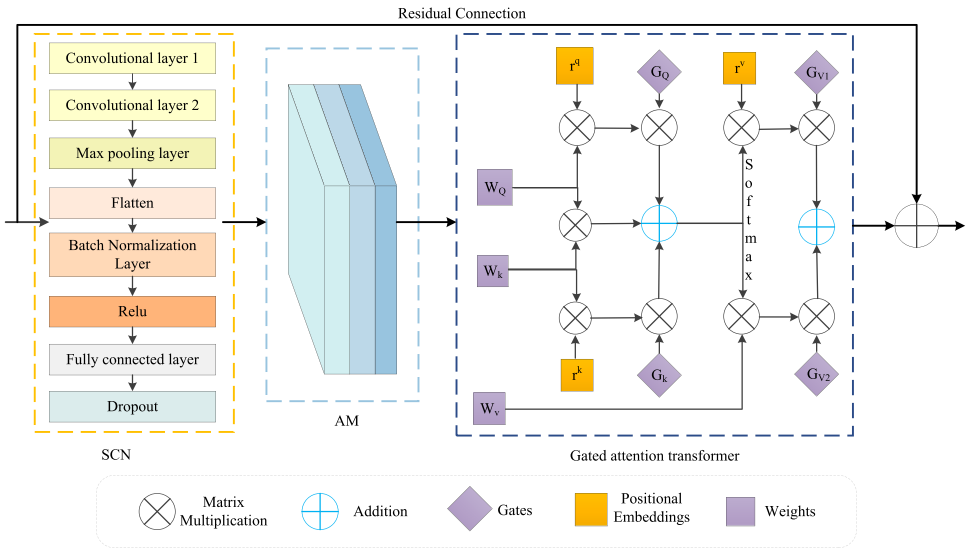

FIGURE 1
AMSCN-GADetector architecture.

events. In order to improve the stability of the training process and effectively alleviate the problem of gradient vanishing or exploding, residual connection is inserted between the spatial convolutional network and the gated attention transformer unit, as well as in the deep structure of the entire network. By directly establishing a fast channel between input and output, stability is provided for the training process of deep network structure, effectively reducing the attenuation phenomenon of gradients when propagating between multiple layers of networks. This enables the model to more flexibly adjust the learning path when capturing the temporal dependence and spatial features of network security events, avoiding the problem of gradient vanishing or exploding caused by increasing network depth. At the same time, it also promotes the smooth flow of information in the network and improves the sensitivity of the model to subtle changes in security anomaly for enterprise management network.

3.2 Spatial convolutional network

In the field of natural language processing, CNNs have shown significant results [29]. While spatial convolutional network fully draws on the advantages of CNNs by constructing double-layer convolutions to deeply explore the dynamic state characteristics of network traffic or log data, achieving efficient extraction and recognition of network anomaly behavior features. In this process, the max pooling layer is used to obtain the most significant anomaly features of enterprise management network security, generate new feature maps and further enhance feature representation capabilities. In spatial convolutional network, convolution layer one and convolution layer two work together to extract key enterprise management network security anomaly features from complex network data using convolution operations. Subsequently, by introducing the max pooling layer and flatten operation, the dimensionality of the data is effectively reduced, making the processing more efficient and reducing computational resource consumption. As a downsampling technique, max pooling is used here to reduce the spatial dimension of network data. This helps to filter out unimportant details and retains features that have a significant impact on the overall structure or pattern of network security anomaly, thereby controlling the risk of overfitting.

In addition, a batch normalization layer [30] is used to alleviate the gradient vanishing problem, and a ReLU activation function is added to introduce nonlinear elements, significantly improving the model's ability to recognize anomaly patterns in complex network. The addition of ReLU activation function enables the model to handle more complex nonlinear relationships, improving the accuracy of detection.

In the feature integration stage, the fully connected layer plays a crucial role. It deeply fuses all the features obtained through convolution and pooling processing, generating higher-level nonlinear feature combinations that can reflect the anomaly behavior characteristics of the network. In order to further improve the generalization ability of the model and prevent overfitting, dropout technique randomly discards a portion of neural connection in each training iteration. This reduces the number of feature detectors, forces the network to learn more stable and generalizes feature representations. This not only helps prevent overfitting of the model to the training data, but also improves the performance of the model in unknown enterprise management network security anomaly detection tasks, enhancing its practical applicability and robustness.

3.3 Attention mechanism

3.3.1 Channel attention

In the graph convolutional network architecture, the input feature dimension is $C \times T \times N$, where C represents the channel dimension, T represents the number of frames and N represents

the number of features within each frame. Given the complexity and continuity of enterprise management network data, in order to effectively retain key security anomaly features, network design usually needs to integrate multiple convolution kernels to extract feature information at different levels of information. However, this process inevitably leads to an increase in the number of output channel, leaving the network with greater challenges in processing information. However, not all channels carry information of equal importance for security anomaly detection. And there are significant differences in the importance of information between channels. To accurately capture and strengthen these key information channels, the channel attention mechanism module is introduced. The core idea of this module is to evaluate the weight of the signal of each channel. The weight level directly reflects the importance of the channel information for the detection of security anomaly. Specifically, the channel with higher weight indicates that the characteristics of the enterprise management network are more related to security anomaly, so more attention should be paid and analysis. The channel attention module consists of two key steps. They are compression and excitation. First, the output feature graph $f_{out} \in R^{C\times T\times N}$ of the convolution of the spatial graph is compressed as input. This step aims to compress the complex spatial feature information, that is the feature distribution of enterprise management network data at different temporal points or temporal periods into a compact and efficient description. So that more accurate weight allocation can be conducted in the future. The specific implementation of compression operation is shown in Equation 1, which focuses on the extraction of global information. It can process and analyze enterprise management network data more efficiently, improve the accuracy and efficiency of security anomaly detection.

$$z = \frac{1}{T\times N}\sum_{i=1}^{T}\sum_{j=1}^{N} u_c(i,j) \tag{1}$$

The compression operation essentially performs a global average pooling. Its core function is to reduce the original high-dimensional $C\times T\times N$ feature map to $C\times 1\times 1$ to refine and integrate the global spatial feature information contained in the enterprise management network data. Specifically, it integrates these complex and multidimensional feature information into the description of each channel, laying a solid foundation for the subsequent feature weight allocation. Subsequently, the incentive phase is further transformed through the enterprise management network data feature map z. The purpose is to dynamically adjust the feature strength based on the importance of channel information for network security anomaly detection tasks. This enables a more accurate identification of potential security threats, and the specific transformation process is shown in Equation 2. This formula can allocate appropriate weights to each channel based on the global feature information obtained during the compression stage, thereby achieving effective screening and optimization of enterprise management network data.

$$s = \sigma(W_2\delta(W_1 z)) \tag{2}$$

The excitation operation specifically contains two fully connected layers. The $w_1 \in R^{c\times\frac{c}{r}}, w_2 \in R^{\frac{c}{r}\times c}$ represents the weight matrix corresponding to the two fully connected layers, δ is the ReLu activation function and σ is the Sigmoid activation function. ReLU can effectively alleviate the gradient disappearance problem and promote the sparsity of the model through its non-negative output characteristics, which helps the model to learn a more stable feature representation. The ReLU activation function plays an important role in the network security anomaly detection, which can effectively alleviate the gradient disappearance problem. It is achieved through its non-negative output feature. This feature promotes the sparsity of the model and enables the model to learn more stable and representative network characteristics. Especially in the face of complex and changeable network attack patterns, ReLU can motivate the model to focus more on capturing the key security anomaly features. While it ignores the unimportant background network information, thus improving the accuracy and efficiency of anomaly detection. On the other hand, the Sigmoid activation function, with its unique S-shaped curve, smoothly maps the input data into the (0,1) interval. This feature is useful in handling certain specific situations in network security anomaly detection, such as the Sigmoid function when detecting subtle changes or soft transitions in network data. These subtle changes may represent potential security threats, and the sensitivity of the Sigmoid function allows the model to identify these threats more accurately.

In the process of realizing the channel attention mechanism, z first uses a fully connected layer to reduce the dimension, aiming to effectively reduce the dimension of the original features to $1\times 1\times C/\tau$. It is used to reduce the complexity of the subsequent calculation, and then optimize the computational efficiency of the whole detection process. Then we apply the ReLU activation function, which significantly enhance the nonlinear expression ability of the network and introduce sparsity through its features, allowing the model to focus more on important channel features and ignore those that contribute less to anomaly detection. This feature is particularly important when handling complex and variable network data because it is able to help the model to identify potential anomaly patterns more effectively. In the second fully connected layer, the original channel dimension of the recovered feature vector is $1\times 1\times C$, ensuring the consistency of the feature map in the dimension. Finally, the Sigmoid activation function is mapped to the continuous interval of 0–1, ensuring the smoothness and continuity of weight allocation. So that the model can adjust more finely the contribution of different channel features when anomaly detection. Figure 2 shows the schematic diagram of network structure after the introduction of channel attention module, which clearly shows how attention weights are combined with other parts of the network to work on feature extraction and representation learning, so as to realize the effective detection of enterprise management network security anomalyities.

3.3.2 Spatial attention

In order to further optimize the network's ability in spatial feature extraction, this paper integrates a spatial attention module. The expression of the spatial attention module is shown in Equation 3, which describes how to dynamically adjust the weights of feature maps based on the importance of spatial positions.

$$s = \sigma(w_s(AvgPool(f_{in}))) \tag{3}$$

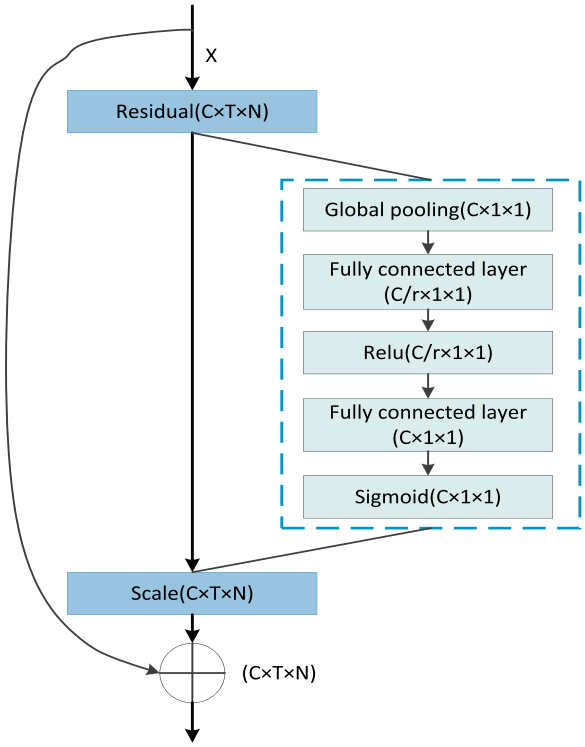

FIGURE 2
Channel attention structure.

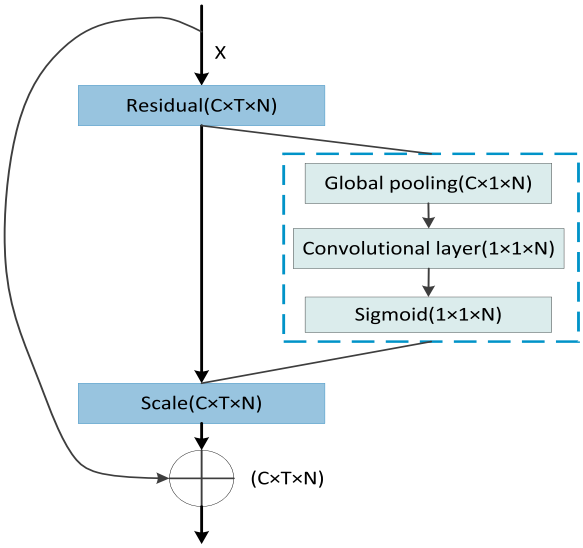

FIGURE 3
Spatial attention structure.

The input is $f_{in} \in R^{C\times T\times N}$, in order to focus on the key feature information of the enterprise management network data at the spatial level, this paper adopts a strategy of pooling the global average *AvgPool* in the temporal dimension. This step effectively compresses the redundant information on the temporal dimension, reducing the dimensionality of the feature graph from the original state to $C\times 1\times N$. Given that key security events or anomaly behaviors are often formed by a few specific network activities or connections, and the spatial distribution of these activities is crucial for exception identification, this paper avoids unnecessary further dimension reduction to retain these key features. The one-dimensional convolution operation w_s has a weight dimension of $1\times C\times K$, the convolution results are subsequently mapped to the interval of 0–one using the Sigmoid activation function to generate the spatial attention weight $s \in R^{1\times 1\times v}$. Finally, to maintain the richness and integrity of the network features, the generated spatial attention weights are combined with the original feature map by means of residual connectivity. This step ensures that the model exploits the spatial attention mechanism without losing any useful information in the original feature graph. Figure 3 illustrates a network structure diagram of the integrated spatial attention module, acting together on the feature extraction and anomaly recognition processes.

Assuming there is a DDoS attack targeting a specific server in an enterprise network. This type of attack can lead to a large number of invalid requests flooding towards the server, resulting in a clear anomalous pattern in the spatial dimension. The spatial attention mechanism can focus on abnormal traffic patterns of specific source or destination IP addresses in network traffic. By assigning higher weights to these key regions, the mechanism can highlight traffic features related to attacks, thereby more effectively extracting spatial features related to abnormal behavior.

3.3.3 Temporal attention

Enterprise management network data typically consists of multiple continuous and interdependent sub events, each of which has varying importance for overall anomaly detection. Therefore, in the task of detecting network security anomaly in enterprise management, it is particularly crucial to allocate temporal attention reasonably. For example, when detecting a potential network attack event, a critical stage of the attack behavior plays a more decisive role in identifying the entire attack behavior compared to the preparatory stage before the attack. In order to focus on the most critical temporal points for anomaly detection, this paper introduces a temporal attention module aimed at highlighting those frames that are most critical for anomaly detection by assigning appropriate weights to each temporal period in the temporal-series network data. This mechanism allows the model to focus more on temporal period that contain important security information, while ignoring temporal period that contribute less to anomaly detection. The temporal attention module has a certain similarity in structure with the spatial attention module, as they both focus on key parts of the input data by calculating weights. However, the temporal attention module focuses on weight allocation in the temporal dimension to adapt to the characteristics of temporal data. The expression of this module is shown in Equation 4, which details the calculation process of temporal attention weights.

$$s = \sigma(w_t(AvgPool(f_{in}))) \tag{4}$$

Specifically, the global average pooling operation *AvgPool* in the temporal attention module is compressed for the spatial dimension to eliminate redundant information in space and thus focus on

feature extraction in the temporal dimension. After a series of calculations, the module outputs a temporal attention weight with a dimension of $1 \times T \times 1$, which directly reflects the importance of each frame in the temporal sequence.

Consider a scenario where a user in the enterprise management network suddenly generates a large number of data access requests during non working hours, which may indicate the presence of malicious software stealing data in the background. This abnormal behavior exhibits clear characteristics in the temporal dimension. The time attention mechanism can identify abnormal patterns in the temporal dimension of network traffic. For example, it can focus on the surge of abnormal requests during non working hours, thereby assigning higher weights to traffic features during these time periods, effectively extracting features related to time abnormal behavior.

3.4 Gated attention transformer

In the gating mechanism, while calculating affinity through self-attention mechanism, we also add a network data position deviation term to make affinity sensitive to network data position information. This bias term is commonly referred to as relative position encoding [31]. These network data location codes can usually be learned through training to enhance the selected features for enterprise management network security data space feature learning, which can help improve the effectiveness of network data detection, effectively reduce training time and enhance detection security performance. Output is calculated using Equation 5 with the help of the self-attention layer.

$$Y_i^a = \sum_{h=1}^{H} \sum_{w=1}^{W} softmax\left(q_{ij}^T k_{hw}\right) v_{hw} \tag{5}$$

where the query $q = W_Q Y_i^{cpstring_2}$, the key $k = W_k Y_i^{cp_2}$ and the value $v = W_v Y_i^{cp_2}$ are both calculated from the input $Y_i^{cp_2}$. And the projection matrices W_Q, W_k, W_v can be learned. As shown in Equation 5, the value v is pooled based on the affinity calculated using softmax ($q^T k$). The attention mechanism can calculate the non-local context with good computational efficiency. It can encode the location bias into the mechanism and can encode the traffic in the input feature mapping [32]. However, it is not always accurate when encoding the network data. And adding them to the respective keys, queries, and value tensors results in a performance degradation when the relative position of the learned network data is not precise enough. Therefore, we propose an improved attention block that can control the effect of positional bias on the coding of nonlocal contextual network data, with the improved attention mechanism shown in Equation 6.

$$Y_i^a = \sum_{w=1}^{W} softmax\left(q_{ij}^T k_{iw} + G_Q q_{ij}^T r_{iw}^q + G_K k_{iw}^T r_{iw}^k\right)\left(G_{v1} v_{iw} + G_{v2} r_{iw}^v\right) \tag{6}$$

where, the improved self-attention formula is closely related to Equation 5, the $r^q, r^k, r^v \in R^{W \times W}$ is axial attention and increases the gating mechanism. The $G_Q, G_k, G_{v1}, G_{v2} \in R$ are learnable parameters. They together form the gating mechanism that controls the effect of the learned relative positional coding on non-local context coding, as shown in Figure 1. Typically, if the relative positional encoding of the network data is accurately learned, the gating mechanism gives it a higher weight than those that are not accurately learned.

4 Analysis of experimental results

4.1 Performance testing evaluation indicators

The experimental environment configuration for this article is Core i7-12700KF, NVIDIA RTX3090, and 64 GB RAM. It builds on the Ubuntu Server operating system, with 32.00 GB of memory and Python as the main programming language. The experimental evaluation utilizes three standard network data security anomaly detection datasets, namely, Bot-IoT [33], CIC-IDS2017 [34] and UNSW-NB15 [35].

In order to verify the effectiveness of enterprise management network security anomaly detection based on AMSCN-GADetector, this paper conducts detailed experimental verification on multiple publicly available and representative datasets. Specifically, Bot-IoT includes various types of attacks such as distributed denial of service (DDoS), denial of service (DoS), operating system and service scanning, keylogging and data breaches, simulating a complex network environment that integrates normal and botnet data. CIC-IDS2017 not only covers various known types of attacks, but also reflects the diversity of network security feature sets and data sources, making it a publicly available dataset that meets real-world standards. This dataset contains 8 files that record five consecutive days of network activity, with a total of 80 features and 2830743 data instances. It also showcases a more diverse range of attack types, making it a resource for evaluating and training network security detection and management models. UNSW-NB15 is generated by IXIA PerfectStorm tool and contains 2540044 traffic data, which are divided into two categories. They are secure and non-secure, including nine types of attacks and one type of normal traffic. We develop 12 algorithms using the ArgusBro-IDS tool and generate 49 features using class labels. These datasets not only provide rich network traffic and behavior data, but also cover various types and scenarios of attacks. This helps to comprehensively evaluate the performance and effectiveness of the model. Meanwhile, the widespread recognition and application of these datasets in the field of network security also ensure the accuracy and credibility of research results.

Given the complex and ever-changing network environment that can lead to significant differences in data, this poses a challenge for training neural network. Therefore, before training the neural network, we normalize the raw data to ensure that the differences in changes between the data were within a controllable range. And we try to keep the values of each data item within the [0,1] interval, thereby accelerating the convergence speed of the program. In terms of experimental evaluation, we use accuracy, precision, recall and F1 as evaluation indicators, the Equations are as follows. Among them, FP is a false positive. In network security anomaly detection, false positives may manifest as mistaking normal network traffic or behavior for attack behavior. FN is a false

negative, which may indicate a failure to detect the true attack behavior in network security anomaly detection. The scarcity of samples may lead to weaker recognition ability of the model for rare attack types during testing, thereby increasing the false negative rate. By comprehensively analyzing and evaluating the detection accuracy, recall, F1, and accuracy of various categories of network security data, we can comprehensively measure the detection performance of models. Especially F1, it is an evaluation index that comprehensively judges accuracy and recall. When there is a discrepancy between accuracy and recall, F1 can provide a more objective evaluation result. The details are presented as Equations 7-10.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FP} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

4.2 Evaluation of network security anomaly detection results

Based on the management network security anomaly detection dataset, we evaluate and select multiple baselines for network security detection comparison experiments, including CNN-SoftMax [36], CNN-LSTM [37], BiGRU [38], AlexNet [39], and GCNN-LSTM [40]. During the training process, the AMSCN-GADetector gradually stabilizes after approximately 20.643 s, with a significant decrease in fluctuation amplitude. This indicates that after efficient learning in the initial stage, the AMSCN-GADetector successfully converges to a relatively stable and high-performance state. The CNN-SoftMax, CNN-LSTM, BiGRU, AlexNet and GCNN-LSTM are 56.54, 50.41, 49.76, 40.97 and 35.36, respectively, indicating that AMSCN-GADetector outperforms these models in terms of convergence speed and performance. In addition, detailed experimental results for all methods are shown in Table 1.

According to Table 1, the experimental results show that the enterprise management network security anomaly detection method based on AMSCN-GADetector performs the best in accuracy, precision, recall and F1. Especially on UNSW-NB15, the accuracy of AMSCN-GADetector is as high as 0.978. And the F1 also reaches 0.974, fully demonstrating its powerful anomaly detection capability. In contrast, CNN-SoftMax performs relatively mediocre on all datasets, with all performance indicators at a low level. Although GCNN-LSTM has shown some competitiveness, its performance on multiple datasets is still slightly inferior to AMSCN-GADetector. In addition, BiGRU and AlexNet also achieve good results on certain datasets, but they are still not comparable to AMSCN-GADetector. These results further validate the effectiveness of attention mechanism in detecting network security anomaly in enterprise management. Figures 4–6 are comparison charts of experimental results on different datasets, where the X-axis represents experimental metrics and the Y-axis represents the corresponding accuracy, precision, recall and F1 of different neural network models.

In Figure 4, compared to other models, AMSCN-GADetector achieves an accuracy of 0.965, which is about 5.8% higher than the second best performing GCNN-LSTM. The recall rate reaches 0.972, which is about 4.7% higher than GCNN-LSTM. It reaches a precision of 0.957, which is approximately 3.6% higher than GCNN-LSTM. The F1 also reaches 0.969, which is about 5.9% higher than GCNN-LSTM. These data fully demonstrate the superiority of AMSCN-GADetector in enterprise management network security anomaly detection tasks, with significantly better performance than other deep learning models, including CNN-SoftMax, CNN-LSTM, BiGRU, AlexNet and GCNN-LSTM.

Based on the same conditions mentioned above, further experiments are conducted on the model. The detailed comparison of the experimental results on CIC-IDS2017 is shown in Figure 5. The AMSCN-GADetector with attention mechanism demonstrates excellent performance. Specifically, the accuracy of AMSCN-GADetector is as high as 0.969, which is about 4.8% higher than the second best performing GCNN-LSTM. In terms of recall, AMSCN-GADetector has reached 0.984, which is about 2.6% higher than GCNN-LSTM. It demonstrates excellent anomaly detection capabilities. These data comparisons not only highlight the advantages of AMSCN-GADetector, but also fully validate the effectiveness of attention mechanism in improving the accuracy and efficiency of network security anomaly detection, providing more reliable support for enterprise management of network security protection.

According to Figure 6, the AMSCN-GADetector with attention mechanism is significantly better than other models. Specifically, the accuracy of AMSCN-GADetector reaches 0.980, which is about 3.1% higher than GCNN-LSTM. And the F1 score is as high as 0.974, which is about 2.5% higher than GCNN-LSTM. These data comparisons not only highlight the outstanding performance of AMSCN-GADetector, but also fully validate the effectiveness of attention mechanism in improving the accuracy of network security anomaly detection. In addition, we use AUC as a key indicator to evaluate the performance of models, which is specifically defined by the ROC curve. ROC is a graph plotted with false positive rate (FPR) as the horizontal axis and true rate (TPR) as the vertical axis. The shape of this curve intuitively reflects the ability of the detection model to distinguish between normal and anomaly network activity.

FPR refers to the proportion of samples that are actually negative classes that the model incorrectly predicts as positive classes. In network security anomaly detection, FPR reflects the false positive rate of the model for normal traffic. A lower FPR means that the model can more accurately distinguish between normal and abnormal traffic, thereby reducing unnecessary alarms and interference. However, reducing FPR may lead to an increase in FNR, meaning that the model may miss some truly anomalous traffic. FNR refers to the proportion of samples that are actually positive classes that the model incorrectly predicts

TABLE 1 Indicator test results on Bot-IoT, CIC-IDS2017 and UNSW-NB15.

Dataset	Model	Accuracy	Recall	Precision	F1
Bot-IoT	CNN-SoftMax	0.810	0.835	0.821	0.858
	CNN-LSTM	0.842	0.865	0.844	0.821
	BiGRU	0.903	0.900	0.898	0.921
	AlexNet	0.885	0.905	0.881	0.889
	GCNN-LSTM	0.907	0.925	0.921	0.920
	AMSCN-GADetector	0.965	0.972	0.957	0.969
CIC-IDS2017	CNN-SoftMax	0.739	0.766	0.712	0.754
	CNN-LSTM	0.874	0.858	0.876	0.871
	BiGRU	0.778	0.819	0.810	0.808
	AlexNet	0.823	0.854	0.853	0.842
	GCNN-LSTM	0.921	0.958	0.940	0.942
	AMSCN-GADetector	0.969	0.984	0.959	0.972
UNSW-NB15	CNN-SoftMax	0.804	0.842	0.871	0.855
	CNN-LSTM	0.827	0.867	0.878	0.885
	BiGRU	0.903	0.939	0.856	0.872
	AlexNet	0.935	0.937	0.928	0.944
	GCNN-LSTM	0.968	0.950	0.949	0.949
	AMSCN-GADetector	0.978	0.969	0.980	0.974

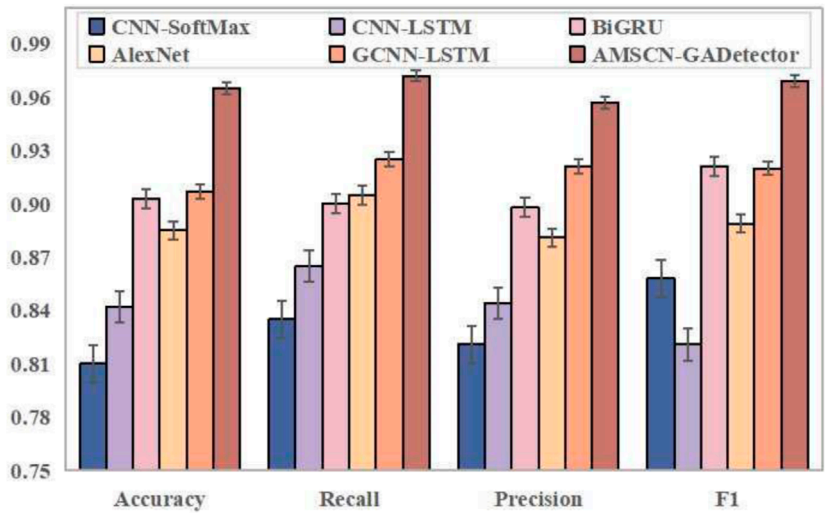

FIGURE 4
Experimental comparison based on Bot-IoT.

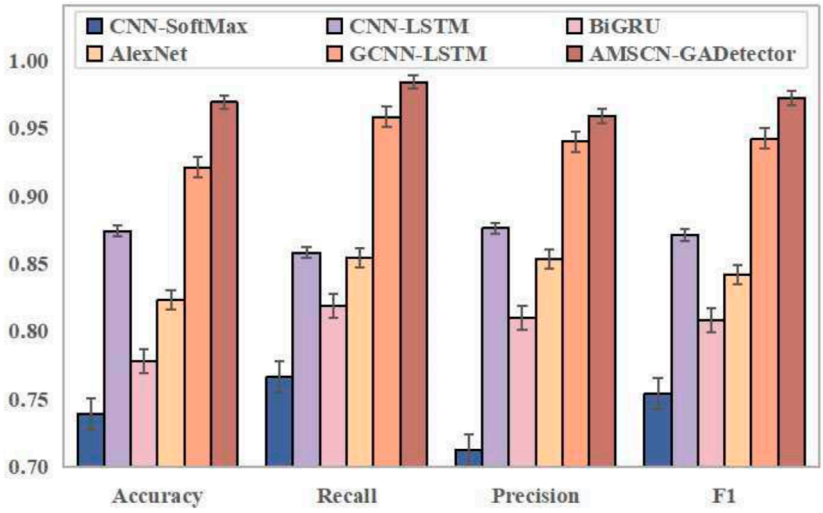

FIGURE 5
Experimental comparison based on CIC-IDS2017.

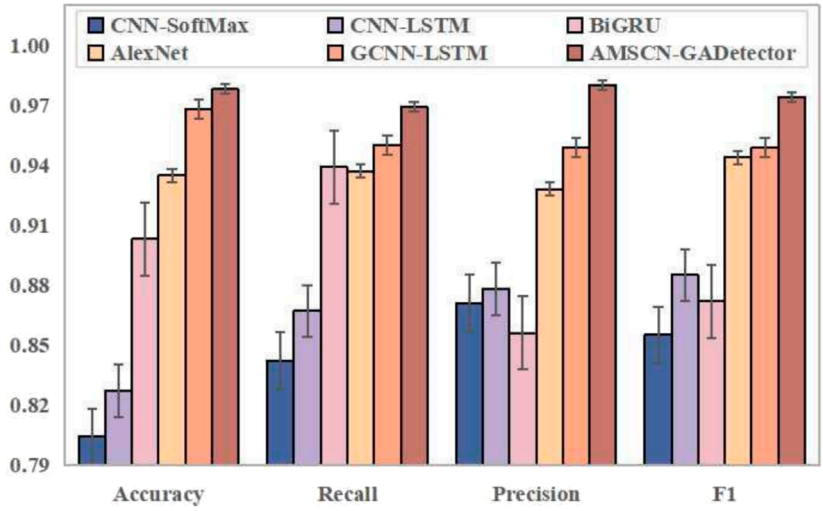

FIGURE 6
Experimental comparison based on UNSW-NB15.

as negative classes. In network security anomaly detection, FNR reflects the missed diagnosis rate of the model for abnormal traffic. A higher FNR means that the model may miss some important abnormal traffic, thereby increasing the risk of network security.

Specifically, the closer the AUC value is to 1.0, the better the performance of the enterprise management network security anomaly detection algorithm, it can more accurately identify potential security threats. On the contrary, when the AUC value reaches 0.5, it means that the performance of the detection model is equivalent to random guessing. That is, the model has no predictive value in judging whether network activity is anomaly. When the AUC value is below 0.5, the detection performance of the model is even inferior to random guessing, which is unacceptable in practical applications. The experimental comparison results based on Bot-IoT, GCNN-LSTM and AMSCN-GADetector are shown in Figure 7.

From Figure 7, it can be seen that AMSCN-GADetector occupies a higher position on the ROC curve, especially in the low FPR region, indicating that the model can more effectively detect true network security anomaly while maintaining a low false alarm rate. The enterprise management network security anomaly detection method based on AMSCN-GADetector first learns rich features through local sub models. Secondly, adding attention mechanism layer during model fusion preserves more key management network data security features, resulting in better detection performance of the trained detection model.

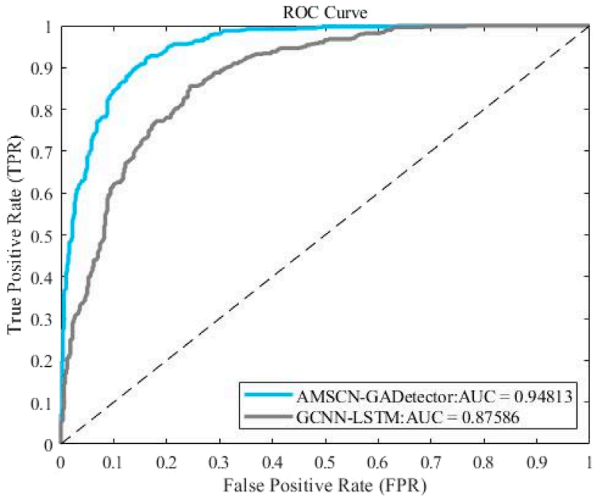

FIGURE 7
Comparison of AUC between GCNN-LSTM and AMSCN-GADetector.

AMSCN-GADetector has powerful feature extraction and non local interaction learning capabilities by integrating spatial convolutional network with gating mechanism. In the actual deployment of real-time data stream processing, it is necessary to allocate computing resources reasonably to ensure the stable operation and efficient processing of the model. At the same time, optimization strategies such as batch processing and asynchronous updates need to be adopted to improve resource utilization and reduce processing costs. For large enterprise management network, distributed deployment strategy can be considered to deploy AMSCN-GADetector on multiple nodes or servers to achieve parallel processing and load balancing, thereby improving the processing capability and response speed of the model.

5 Conclusion

With the rapid development of information technology, as a key infrastructure for promoting enterprise operation and development, enterprise management network is facing increasingly severe network security threats. Therefore, this article proposes AMSCN-GADetector for enterprise management network security anomaly detection, which is based on deep learning for complex network data security identification, anomaly detection and intelligent analysis. AMSCN-GADetector consists of three core modules, namely, attention mechanism, gated attention transformer and spatial convolutional network. This model deeply integrates global contextual information and can efficiently and comprehensively extract global deep spatial features from enterprise management network. More importantly, by introducing attention mechanism, AMSCN-GADetector can dynamically adjust the weight of dependency relationships between enterprise management network data, effectively improving the model's ability to capture key information of security anomaly data, thereby further improving the accuracy of detection. Through experimental and comparative analysis, it is concluded that AMSCN-GADetector exhibits significant advantages in performance indicators such as accuracy, precision, recall and F1. Its network security detection and recognition performance are significantly better than other baseline models, providing strong support for improving enterprise management network security protection capabilities and ensuring operational security.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

ZY: Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing–original draft. YZ: Formal Analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Pereira T, Barreto L, Amaral A. Network and information security challenges within Industry 4.0 paradigm. *Proced manufacturing* (2017) 13:1253–60. doi:10.1016/j.promfg.2017.09.047

2. Ahmed M, Mahmood AN, Islam MR. A survey of anomaly detection techniques in financial domain. *Future Generation Computer Syst* (2016) 55:278–88. doi:10.1016/j.future.2015.01.001

3. Miao J, Wang Z, Ning X, Xiao N, Cai W, Liu R. Practical and secure multifactor authentication protocol for autonomous vehicles in 5G. *Softw Pract Experience* (2022) 54:1852–69. doi:10.1002/spe.3087

4. Bringhenti D, Marchetto G, Sisto R, Valenza F. Automation for network security configuration: state of the art and research trends. *ACM Comput Surv* (2023) 56(3):1–37. doi:10.1145/3616401

5. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. *ACM Comput Surv* (2009) 41(3):1–58. doi:10.1145/1541880.1541882

6. Xu H. The design of computer network virus defense system based on data mining technology. *J Phys Conf Ser IOP Publishing* (2022) 2173(1):012072. doi:10.1088/1742-6596/2173/1/012072

7. Bondan L, Wauter T, Volckaert B, De Turck F, Granville LZ. NFV anomaly detection: case study through a security module. *IEEE Commun Mag* (2022) 60(2):18–24. doi:10.1109/mcom.001.2100408

8. Miao J, Wang Z, Wang M, Garg S, Hossain MS, Rodrigues JJ. Secure and efficient communication approaches for Industry 5.0 in edge computing. *Computer Networks* (2024) 242:110244. doi:10.1016/j.comnet.2024.110244

9. Wang L, Wang Z. Research on the current situation and technical exploration of network security in power monitoring system. *Trans Computer Sci Intell Syst Res* (2023) 1:134–8. doi:10.62051/96h2c650

10. Xiao N, Wang Z, Sun X, Miao J. A novel blockchain-based digital forensics framework for preserving evidence and enabling investigation in industrial Internet of Things. *Alexandria Eng J* (2024) 86:631–43. doi:10.1016/j.aej.2023.12.021

11. Sejnowski TJ. The unreasonable effectiveness of deep learning in artificial intelligence. *Proc Natl Acad Sci* (2020) 117(48):30033–8. doi:10.1073/pnas.1907373117

12. Sarker IH. Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective. *SN Computer Sci* (2021) 2(3):154. doi:10.1007/s42979-021-00535-6

13. Soydaner D. Attention mechanism in neural networks: where it comes and where it goes. *Neural Comput Appl* (2022) 34(16):13371–85. doi:10.1007/s00521-022-07366-3

14. Du M., Li F., Zheng G., Srikumar V. Deeplog: Anomaly detection and diagnosis from system logs through deep learning. In Proceedings of the 2017 ACM SIGSAC conference on computer and communications security. (2017). 1285–1298. doi:10.1145/3133956.3134015

15. Javaid A., Niyaz Q., Sun W., Alam M. A A deep learning approach for network intrusion detection system. In Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS). (2016). 21–6. doi:10.4108/eai.3-12-2015.2262516

16. Shaikh RA, Shashikala SV. An Autoencoder and LSTM based Intrusion Detection approach against Denial of service attacks. In: *2019 1st international conference on advances in information technology (ICAIT)*. IEEE (2019). p. 406–10.

17. Su T, Sun H, Zhu J, Wang S, Li Y. BAT: deep learning methods on network intrusion detection using NSL-KDD dataset. *IEEE Access* (2020) 8:29575–85. doi:10.1109/access.2020.2972627

18. Derhab A, Aldweesh A, Emam AZ, Khan FA. Intrusion detection system for internet of things based on temporal convolution neural network and efficient feature engineering. *Wireless Commun Mobile Comput* (2020) 2020:1–16. doi:10.1155/2020/6689134

19. El-Sayed O. A., Fawzy S. K., Tolba S. H., Salem R. S., Hassan Y. S., Ahmed A. M., et al. Deep learning framework for accurate network intrusion detection in ITSs. In 2021 International Conference on Microelectronics (ICM). (2021). 212–5. IEEE. doi:10.1109/ICM52667.2021.9664897

20. Ma H., Cao J., Mi B., Huang D., Liu Y., Li S. A GRU-Based Lightweight System for CAN Intrusion Detection in Real Time. Security and Communication Networks. (2022)(1), 5827056. doi:10.1155/2022/5827056

21. Van Huong P, Hung DV. Intrusion detection in IoT systems based on deep learning using convolutional neural network. In: *2019 6th NAFOSTED conference on information and computer science (NICS)*. IEEE (2019). p. 448–53.

22. Andresini G, Appice A, Caforio FP, Malerba D, Vessio G. ROULETTE: a neural attention multi-output model for explainable network intrusion detection. *Expert Syst Appl* (2022) 201:117144. doi:10.1016/j.eswa.2022.117144

23. Cao B., Li C., Sun J., Song Y. IoT intrusion detection technology based on Deep learning. In 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA). (2022). 284–9. doi:10.1109/CVIDLICCEA56201.2022.9825291

24. Lee JH, Park KH. AE-CGAN model based high performance network intrusion detection system. *Appl Sci* (2019) 9(20):4221. doi:10.3390/app9204221

25. Yao R, Wang N, Chen P, Ma D, Sheng X. A CNN-transformer hybrid approach for an intrusion detection system in advanced metering infrastructure. *Multemporaldia Tools Appl* (2023) 82(13):19463–86. doi:10.1007/s11042-022-14121-2

26. Elsayed N., Zaghloul Z. S., Azumah S. W., Li C. Intrusion detection system in smart home network using bidirectional LSTM and convolutional neural networks hybrid model. In 2021 IEEE international midwest symposium on circuits and systems (MWSCAS). (2021). 55–8. IEEE. doi:10.1109/MWSCAS47672.2021.9531683

27. Shone N, Ngoc TN, Phai VD, Shi Q. A deep learning approach to network intrusion detection. *IEEE Trans emerging Top Comput intelligence* (2018) 2(1):41–50. doi:10.1109/tetci.2017.2772792

28. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. *Neurocomputing* (2021) 452:48–62. doi:10.1016/j.neucom.2021.03.091

29. Zhou YT. Natural language processing with improved deep learning neural networks. *Scientific Programming* (2022) 2022(1):1–8. doi:10.1155/2022/6028693

30. Bjorck N., Gomes C. P., Selman B., Weinberger K. Q. Understanding batch normalization. Advances in neural information processing systems, (2018). 31.

31. Zheng W, Gong G, Tian J, Lu S, Wang R, Yin Z, et al. Design of a modified transformer architecture based on relative position coding. *Int J Comput Intelligence Syst* (2023) 16(1):168. doi:10.1007/s44196-023-00345-z

32. Wen Y, Xu P, Li Z, Xu W, Wang X. RPConvformer: a novel Transformer-based deep neural networks for traffic flow prediction. *Expert Syst Appl* (2023) 218:119587. doi:10.1016/j.eswa.2023.119587

33. Peterson JM, Leevy JL, Khoshgoftaar TM. A review and analysis of the bot-iot dataset. In: *2021 IEEE international conference on service-oriented system engineering (SOSE)*. IEEE (2021). p. 20–7.

34. Rosay A., Cheval E., Carlier F., Leroux P. Network intrusion detection: A comprehensive analysis of CIC-IDS2017. In 8th international conference on information systems security and privacy. SCITEPRESS-Science and Technology Publications (2022). 25–36. doi:10.5220/0000157000003120

35. Meftah S., Rachidi T., Assem N. Network based intrusion detection using the UNSW-NB15 dataset. International Journal of Computing and Digital Systems, 2019, 8(5):478–87. doi:10.12785/ijcds/080505

36. Siripibal N, Supratid S, Sudprasert C. A comparative study of object recognition techniques: softmax, linear and quadratic discriminant analysis based on convolutional neural network feature extraction. In: *Proceedings of the 2019 international conference on management science and industrial engineering* (2019). p. 209–14.

37. Zha W, Liu Y, Wan Y, Luo R, Yang S, et al. Forecasting monthly gas field production based on the CNN-LSTM model. Energy, 2022, 260:124889. doi:10.1016/j.energy.2022.124889

38. Lin X, Quan Z, Wang ZJ, Huang H, Zeng X. A novel molecular representation with BiGRU neural networks for learning atom. *Brief Bioinformatics* (2020) 21(6):2099–111. doi:10.1093/bib/bbz125

39. Yu W., Yang K., Bai Y.,**ao T., Yao H., Rui Y. Visualizing and comparing AlexNet and VGG using deconvolutional layers. In Proceedings of the 33 rd International Conference on Machine Learning. (2016), (3), 43–76.

40. Wu Y, Zhu X, Huang Q, Zhang Y, Evans J, He S. Predicting the quality of tangerines using the GCNN-LSTM-AT network based on vis–NIR spectroscopy. *Appl Sci* (2023) 13(14):8221. doi:10.3390/app13148221