Received 21 April 2022, accepted 27 May 2022, date of publication 10 June 2022, date of current version 27 June 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3182010

FETTrans: Analysis of Compound Interference Identification Based on Bidirectional Dynamic Feature Adaptation of Improved Transformer

BORAN LI[1], LEI ZHANG[1], JINGWEI DAI[1], JIAN MA[1], AND ZHIMIN ZHAO[2]
[1]Institute of Wireless and Terminal Technology, China Mobile Research Institute, Beijing 100032, China
[2]Wuxiang Digital Information Center, China Mobile Guangxi Company, Guangxi 530000, China

Corresponding author: Boran Li (liboran@chinamobile.com)

ABSTRACT With the development of 5G network construction and the influx of users, the anomaly detection of base stations become more onerous. The types of uplink interference from base stations have grown significantly, and they will be affected by more than one type. It will be difficult to identify the uplink compound interference for the network operator. Meanwhile, the same interference type has flexible frequency location and variable frequency bandwidth. For multiple interference source cases, the classification accuracy and generalization ability of traditional algorithms cannot meet the needs of the actual production environment. Therefore, this paper proposes the FETTrans network to transform the compound interference identification task into a multi-label classification in natural language processing. Inspired by the attention mechanism in Transformer and combined with the pattern recognition task, the correlation between the constructed output vector and input features is obtained through the attention mechanism. On this basis, the network transforms the unidirectional static projection method into a bidirectional dynamic adaptation method, which effectively improves the identification accuracy of the variable compound interference frequency band. After verification by different test sets collected from the current network, the classification accuracy has steadily increased by more than 15%, and the mAP has reached 92%. The FETTrans network, as an end-to-end algorithm, can be adapted to another fault diagnosis.

INDEX TERMS Anomaly detection, 5G network, multi-label classification, transformer, attention mechanism, fault diagnosis.

I. INTRODUCTION

With the development of the communication network from 4G to 5G, the widespread deployment of base stations, and the access of a large number of new subscribers, the current network faces more serious uplink interference problems. The quality of network optimization directly affects the subscribers' experience. Therefore, optimizing compound interference identification is beneficial to improving network service. China Mobile Communications Corporation (CMCC) has the largest number of base stations and subscribers in China, so we perform compound interference identification based on its current network data.

The 5G base station of CMCC can receive the carrier signal of 2515-2675MHz.The signal intensity in this frequency band is reflected by the power (dBm) of 273 physical resource blocks (PRB). Different interference signals will form different frequency band characteristics. We performed interference type identification by analyzing different data distributions. With the continuous development, the types of interference received by the base station are also increasing. In addition, the superposition of interfering signals and the ever-changing user profiles will greatly increase the difficulty of operation and maintenance. This is because the interference frequency band is flexible and the composite signal waveforms are superimposed. Multiple classifications of PRB data are required due to the superposition of multiple interference powers. Therefore, to improve the operation and maintenance efficiency and reduce the operation cost, it is necessary to design an end-to-end algorithm. And it has a good generalization ability for data in different regions and periods.

This article identifies ten common types of interference in 5G at this stage. The generation principle of each type of

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

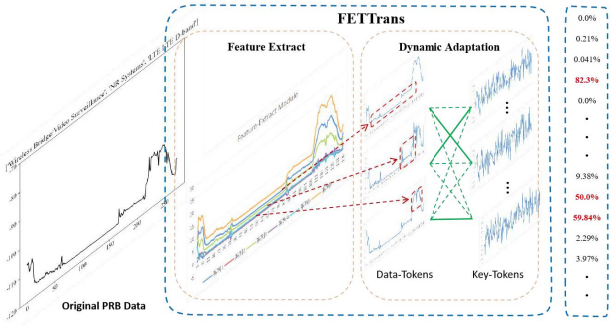


FIGURE 1. Design diagram.

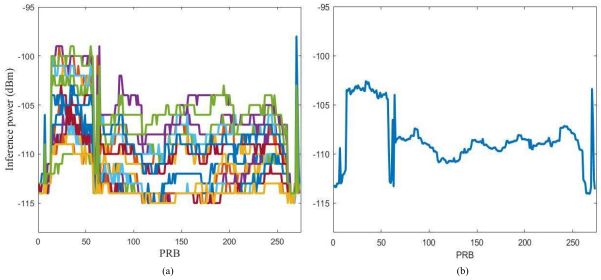


FIGURE 2. PRB dBm every 15 minutes (a) and average PRB dBm (b).

interference, frequency band characteristics, and their visualization images are in Appendix. There are several challenges with current network data that can affect the classification performance of the algorithm. (a) From the perspective of the overall data distribution, different interference sources generate complex and diverse data. Since the communication frequency band is fixed, the interference frequency band will be divided into full-band interference and narrowband interference. The narrowband interference is also divided into flexible frequency point position and fixed frequency point position. (b) There will be similar waveforms in the data distribution of different interference types. The frequency band characteristics between "*Wireless Bridge / Video Surveillance Interference*" and "*Router Interference*" are only different in certain PRB values. (c) The waveform of the compound interference after the signal is superimposed will be confused with other interference types. For example, the superposition of "*NR system Interference*" and "*Pseudo Base Station Interference*" easily form a waveform similar to "*LTE D-band Interference*". In addition, some interference types account for a small proportion of the current network, resulting in training data imbalance.

In order to solve the above problems, this paper proposes the FETTrans algorithm network. Figure 1 is a design diagram based on the network. The network achieves a bidirectional dynamic adaptation process, which greatly improves the recognition accuracy and generalization ability. The leftmost data in Fig. 1 is the original data. Based on the data, the Feature Extract Module performs feature extraction and dimension expansion. The output result, Data-Token, is the input of the FETTrans network. In the FETTrans network, the feature adaptation work is performed by the Key-Tokens of the constructed output subspace and the input Data-Tokens, as shown in the green line in Fig. 1. Through the adaptation of input and output, the bidirectional dynamic training process is realized, and the specific implementation process is described in detail in Section III.B. Finally, FETTrans outputs the compound interference result.

The structure of this paper is divided into six parts. Section II introduces the current technology. Section III introduces the Feature Extract module and the architecture of the FETTrans network respectively. Section IV introduces the datasets and results comparison. Section V concludes this paper. Appendix introduces the principle of interference generation and frequency band characteristics.

II. RELATED WORK

Interference identification in wireless communication systems is generally based on composite time-frequency domain analysis. The FETTrans network proposed in this paper performs compound interference identification based on the PRB data of the base station (PRB represents the physical resources in the time-frequency domain [1]). Figure 2 is a visualization of PRB data. Fig. 2 (a) is all the data collected by the base station every 15 minutes in one day, the abscissa represents the PRB number, and the ordinate represents the signal intensity in the frequency band. So each curve represents the time-frequency domain signal intensity of all signal sources aliasing received by the base station. According to the actual situation, the operator will perform compound interference identification based on the average PRB value over a period of time. Fig.2 (b) is the average signal intensity in the time domain. As shown in Fig. 2, the following conclusions can be drawn:

- Serialized data: The average signal intensity in the time domain forms serial data in the frequency domain;
- Information density: The data contains relatively high information density, which is manifested in the presence of indistinguishable frequency band characteristics while the overall data specificity is high;
- Compound concurrency: The compound interference signals are generally superimposed linearly, and the distribution of data features is prone to confusion.

In recent years, the research of wireless interference identification algorithms has gradually focused on deep learning algorithms. Through the research on three narrow-band interference sources such as Bluetooth, Zigbee and WiFi, the designed CNN model improves the recognition accuracy and convergence speed [2]. For multi-label classification tasks, 90% accuracy is achieved with 1D convolutional neural networks [3]. According to the WLAN technology communication in the factory environment, the interference identification is realized according to the spectral characteristics by training CNN [4]. For the task of interference identification on IoT platforms, COTS hardware is used to

capture signal physical properties and lightweight supervised learning for identification [5]. According to our research, many areas of the industry need to process sensor data for anomaly detection or fault diagnosis. For example, multi-classification failure diagnosis of the motor bearing diagnosis case [6], concurrent failure diagnosis of rotating machinery [7]. The sensor data has similarities to the distribution characteristics above. Most researchers or engineers generally use the following three methods for their field:

- Preprocess data through signal decomposition and use improved machine learning algorithms to solve classification problems. Compound classification is solved by multiple binary classification voting [8], or by training multiple models, such as XGBoost [9];
- Multi-layer neural network is used for feature extraction, such as ELM [10] or LSTM [11];
- Generate pictures through visualization techniques, and use CNN, ResNet [12] and other image recognition algorithms to complete the classification task.

The above algorithms are more effective for solving multi-classification tasks, and the accuracy is less than 90%. Generally, the multi-label problem is solved by transforming into multiple single-class problems [13]. And in the actual production environment, we found that there are different degrees of overfitting for solving multi-label classification tasks. This leads to compound interference identification in different regions, or after a period of time, the classification accuracy of the algorithm will drop significantly. In addition to the decrease in the accuracy of the model due to changes in user portraits and interference sources, it is difficult for the model to learn key features from a large and unbalanced multi-label training set. It is difficult to collect some interference data in the current network, and the workload of calibrating the data set is huge. So we are committed to building an end-to-end algorithmic network that solves the multi-class compound output. According to data characteristics, we found that this is similar to many tasks in NLP. It is to translate different data features in serialized data into different tokens in NLP. In this process, we were inspired by the Transformer [14] algorithm, in which the multi-head attention mechanism in Transformer is very necessary for the feature extraction work of processing sequence signals. Eq. (1) is the implementation of the attention mechanism.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V, \tag{1}$$

Transformers have achieved state-of-the-art results in a wide range of natural language tasks including generative language modeling [15], [16] and discriminative language understanding [17]. This success is partly due to the self-attention component, which enables the network to capture contextual information from the entire sequence [18]. The seq2seq model can effectively solve the relationship between multiple inputs and multiple outputs by encoding the input and output. Therefore, we hope that the attention mechanism can effectively adapt the relationship between input features and output encoding. In addition to this, we noticed that the multi-head attention in the Transformer has better generalization performance. Although this paper is from the perspective of images, it theoretically analyzes the role of multi-head attention and draws conclusions [19]. In fact, the performance and potential of Transformer are obvious to all, and the adaptation to large data sets is also better. Therefore, we design a network structure based on the above advantages to transform the multi-label classification from ''unidirectional static projection'' to ''bidirectional dynamic adaptation'' method.

III. FETTrans NETWORK

A. FEATURE EXTRACT MODULE

1) CONV-EXTRACT

The input signal is obviously difficult to transform a classification task into the seq2seq task. We borrowed the ideas of the Inception [20] algorithm and designed a multi-channel and multi-size 1D convolution kernel for dimension expansion and filtering at different scales. The design diagram of the Conv-Extract module (CoE) is shown in Fig. 3.

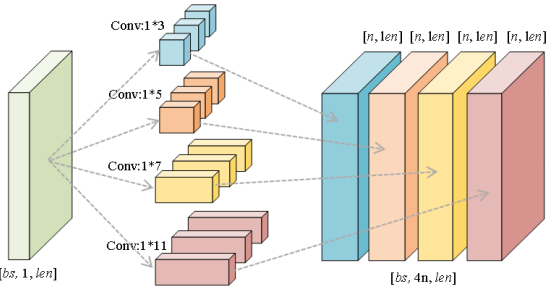


FIGURE 3. Conv-extract module.

In Fig. 3, *bs* represents the batch size, *len* represents the sequence length of the data, and *n* represents the number of convolution kernel channels. The input is one-dimensional origin data. Since the influence bandwidth is different, the CoE module contains 4 different sizes of one-dimensional convolution kernels: 1*3, 1*5, 1*7, 1*11. The input data preliminarily completes the feature extraction through it, and the data is expanded to *4n* dimensions. The output multi-dimensional data expresses complex superimposed signals from different scales, which improves the expressive ability of the model. This is similar to signal decomposition in the field of fault diagnosis, but the CoE is more inclined to complete the extended dimension task.

Based on the current network data, we visualize the data generated by the CoE. Fig. 4(a)-(d) are the results in Fig. 1. From the verification results, it effectively extracts critical frequency band characteristics. And the more obvious the data features, the greater the extracted signal intensity. The most obvious in the original signal is the ''*Wireless Bridge / Video Surveillance Interference*''. The orange curve of the CoE output in Fig. 4 (b) has more obvious frequency band characteristics, and the signal intensity is also significantly

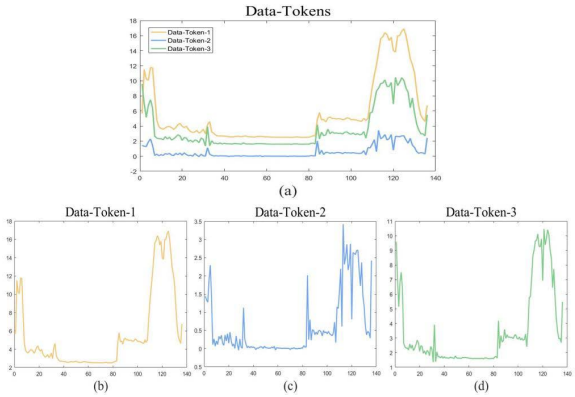


FIGURE 4. Output of the CoE module.

higher than other curves. For "*LTE D-band Interference*" and "*NR system Interference*", their frequency band characteristics are relatively insignificant in this sample, so the signal intensity is not high. But compared to Fig. 4 (b), Fig. 4 (c) has more obvious characteristics of "*LTE D-band Interference*" and Fig. 4 (d) has more obvious characteristics of "*NR system Interference*" and "*Wireless Bridge / Video Surveillance Interference*".

2) FEATURE PROJECT

Although the multi-channel convolution kernel can effectively expand the signal dimension and filtering effect, its expressive ability is still lacking. The Feature Project module projects multi-dimensional signal features into the input subspace through four layers of fully connected layers and ReLU activation function. It is able to further improve the expressiveness of the network. And the Feature Project module does not scale the data dimension. It is just used to extract abstract features and output the Data-Tokens matrix. Fig. 5 is the Feature Project module design diagram.

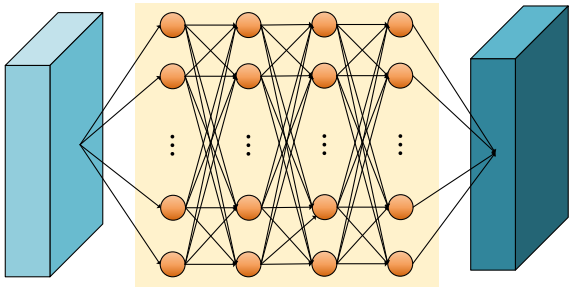


FIGURE 5. Feature project module.

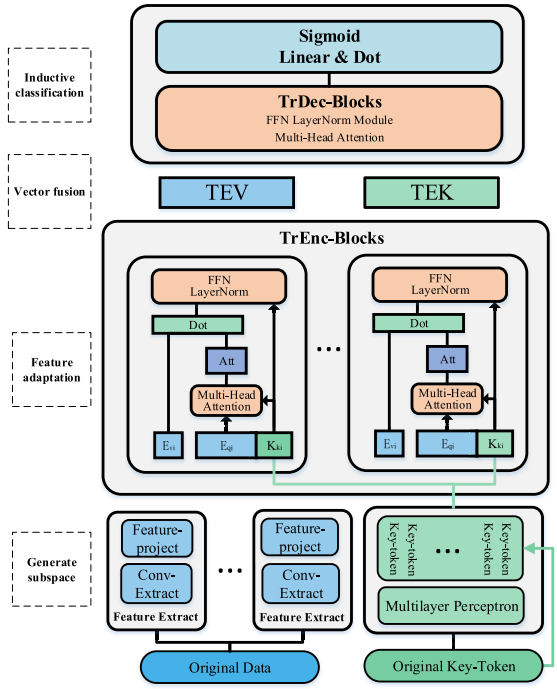


FIGURE 6. FETTrans network.

B. FEATURE TOKENS TRANSLATE

This section describes how to transform compound interference into a natural language processing task. The input data is multi-feature input data containing high information density. Multiple data features correspond to multiple signal interferences, which is equivalent to multiple inputs corresponding to multiple outputs. Since compound interference identification is essentially a pattern recognition task, there is no contextual semantic information in either input or output. For the same interference type, the data characteristic distribution is not unique. It is different from a word vector in natural language processing. Therefore, the input information is relatively ambiguous for the "translation" task, and there is also signal superposition. Of course, the connection between contexts in natural language processing and the memory problem of long sentence patterns do not need to be considered in the compound interference recognition task.

For the above feasibility, it mainly depends on the Transformer algorithm structure that can process the input and output data in parallel, rather than the recurrent processing of the RNN network. Inspired by the Transformer, the output of the Feature Extract (FEt) module is regarded as Data-Tokens. The output categories are then represented by constructing encoding vectors (Key-Tokens) of the output subspace. Through explicit coding, Key-Tokens can be trained with data features through the network, which helps deal with interference signals of different frequency bandwidths. According to the above analysis, we strengthen the relationship between input features and output categories and weaken the contextual semantic relationship. So the FETTrans network implements a new solution. First, in TrEnc-Blocks (Encoder), Data-Tokens and Key-Tokens are used to complete the feature bidirectional adaptation work through the multi-head attention mechanism. This part mainly strengthens the adaptation performance of a single category. Then the input and output after feature adaptation are extracted separately. After the overall input data and overall output data are input into

TrDec-Blocks (Decoder), their correlation degree is obtained again through the multi-head attention mechanism. Finally, the classification result is obtained through the fully connected layer of the Sigmoid activation function. Then through the multi-hot label and MSE loss function, the multi-label classification task is solved. In this way, we realize the process of transforming unidirectional static projection into bidirectional dynamic adaptation. Based on the above analysis, we design the network structure, as shown in Fig. 6.

Fig. 6 is divided into four parts from bottom to top: Generate subspace, Feature adaptation, Vector fusion, Inductive classification.

- Generate subspace: Assume that the number of classification types is n, and the length of the data sequence is m. The original data forms an n-dimensional input subspace through n FEt modules, and each FEt module expands the data dimension (k-dimensional) through a multi-channel convolution kernel to obtain **E** matrix of $[n, k, m]$ shape. We construct n Original Key-Tokens (all-ones vectors) in the output subspace. After each dimension Original Key-Token is input to the corresponding fully connected layer, the residual connection with the input is performed to obtain the $[n, 1, m]$ shape Key-Token output encoding matrix, which is denoted as **K** matrix;
- Feature adaptation: TrEnc-Blocks implements feature adaptation between input features and output encodings in subspaces.
  - First, through different fully connected feedforward layers, $\boldsymbol{E}_i$ (the ith of **E**) is linearly projected to obtain $\boldsymbol{E}_{qi}$ and $\boldsymbol{E}_{vi}$, and $\boldsymbol{K}_i$ (the ith of **K**) is projected to $\boldsymbol{K}_{ki}$.
  - Then, $\boldsymbol{E}_{qi}$, $\boldsymbol{E}_{vi}$ and $\boldsymbol{K}_{ki}$ are respectively input into TrEnc-Block, where Multi-Head Attention and FFN LayerNorm are similar to the Encoder Blocks of Transformer. The difference is: First, **Att** is not the result of self-attention, but the degree of correlation between the k-dimensional feature of $\boldsymbol{E}_{qi}$ and the i-th dimensional Key-Token. Second, the input of the FFN module is obtained by concatenating Kki and dot product ($\boldsymbol{E}_{vi}$ and **Att**), and the output is denoted as $\boldsymbol{O}_v$. Third, there are no residual connections in the FFN LayerNorm.
- Vector fusion: Since the FFN module does not change the input shape, the shape of $\boldsymbol{O}_v$ is a $(n,2,m)$ matrix. By recombining $\boldsymbol{O}_v$, the **TEV** matrix and the **TEK** matrix are obtained.

$$\begin{aligned} \mathbf{TEV} &= \mathrm{Concat}_{i=1}^{n}(O_v[i, 0, m]) \\ \mathbf{TEK} &= \mathrm{Concat}_{i=1}^{n}(O_v[i, 1, m]) \end{aligned} \tag{2}$$

  where "0" represents the first dimension of $\boldsymbol{O}_v$, and "1" represents the second dimension of $\boldsymbol{O}_v$.
- Inductive classification: The **TEV** and **TEK** matrices are the input of TrDec-Blocks, respectively. There is no masked self-attention mechanism and position encoding information in Decoder Blocks in TrDec-Blocks. After the output of TrDec-Blocks is subjected to linear projection and dot product operation, the result is subjected to Sigmoid function to obtain the classification result.

$$\begin{aligned} y_c = \mathrm{Sigmoid}(&\mathrm{ReLU}(W_{\mathrm{TEV}}x_{\mathrm{TEV}} + b_{\mathrm{TEV}}) \cdot \\ &\mathrm{ReLU}(W_{\mathrm{TEV}}x_{\ \mathrm{TEV}\ } + b_{\mathrm{TEV}})) \end{aligned} \tag{3}$$

Fig. 7 is a heat map of y_c, and the input data is the data sample in Fig. 1.

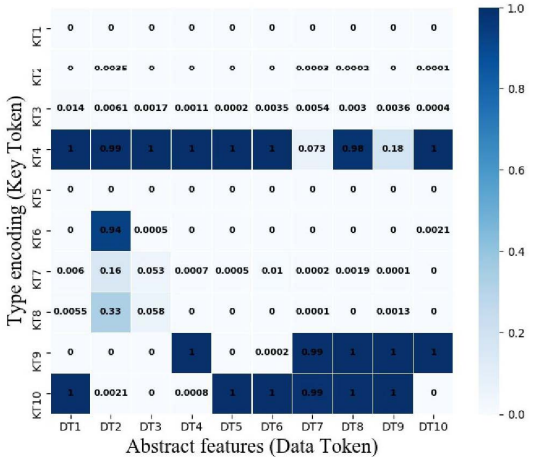


FIGURE 7. Heat map of y_c.

The abscissa is the abstract feature of the input subspace, where $\mathbf{DT}_i$ represents the i-th Data Token of the input. The ordinate is the Key-Token of the output subspace. In addition to representing the i-th Key Token, $\mathbf{KT}_i$ also represents the interference type. The correspondence is in Tab. 1. y_c is actually the correlation between the input **TEV** and **TEK**. By averaging the vertical axis direction, the final classification result can be obtained. The results are shown in Tab. 2. The values in the table are percentages.

We successfully transform the compound interference classification task into a natural language processing task based on the FETTrans network. The most important thing is to realize a bidirectional dynamic feature adaptation work by constructing the Key-Token of the output subspace. The difference between FETTrans and Transformer is as follows:

- FETTrans can achieve fully parallel processing since TrDec-Blocks has no masked self-attention mechanism;
- The implementation of feature adaptation depends on the input $\boldsymbol{K}_{ki}$ vector;
- The input of TrDec-Blocks is not directly obtained from the embedding layer in NLP;
- The output result of TrDec-Blocks obtains multi-classification results through the Sigmoid function, and the loss function is MSE.

Finally, the FETTrans network can complete the end-to-end multi-label classification task by changing the number of identified types.

TABLE 1. The effective candidate patterns and their corresponding parameters.

	Type of interference
$\mathbf{KT}_1$	Intelligent Street Lamp Interference
$\mathbf{KT}_2$	China Telecom LTE 800MHz Band Interference
$\mathbf{KT}_3$	MMDS signal Interference
$\mathbf{KT}_4$	Wireless Bridge / Video Surveillance Interference
$\mathbf{KT}_5$	Jammer Interference
$\mathbf{KT}_6$	Router Interference
$\mathbf{KT}_7$	Pseudo Base Station Interference
$\mathbf{KT}_8$	Atmospheric Dust Interference
$\mathbf{KT}_9$	NR Systems Interference
$\mathbf{KT}_{10}$	LTE D-band Interference

TABLE 2. Probability of each interference type.

	$\mathbf{KT}_1$	$\mathbf{KT}_2$	$\mathbf{KT}_3$	$\mathbf{KT}_4$	$\mathbf{KT}_5$	$\mathbf{KT}_6$	$\mathbf{KT}_7$	$\mathbf{KT}_8$	$\mathbf{KT}_9$	$\mathbf{KT}_{10}$
Prob(%)	0.0	0.2	0.39	**82.3**	0.0	9.4	2.3	4.0	**50.0**	**59.8**

IV. RESULT AND COMPARISON

A. DATASET DESCRIPTION

The validation of this scheme is based on datasets provided by two provinces in China. The data set of province A was collected in two periods. In June 2021, we collected a total of 10,596 samples in Province A. The dataset is divided by 20%, with 8476 samples in the training set and 2093 samples in the test set (T_A1). Then, the test set (T_A2) of Province A was collected in December 2021. During this period, more 5G users joined the network, which will lead to an increase in the noise floor. There are a total of 686 calibration data. The data set of province B (T_B) consists of 1033 samples collected in December 2021, which is a cross-province and cross-time test set. Since the calibration and labeling work needs to be determined by professionals through unified negotiation, and the wrong labels need to be continuously reviewed, there are few valid datasets. In the current network, some interference types appear less frequently, resulting in unbalanced labels. The following table shows the number of samples of each interference type in the training set.

Tab. 3 is the classification and statistical results of interference types. ''All type sample'' indicates the number of interference types in the overall dataset. ''Single type sample'' represents the number of samples with a single classification label. And ''Compound Type sample'' is the number of samples of this type in the compound classification label.

The feature distributions of $\mathbf{KT}_4$(*Wireless Bridge / Video Surveillance Interference*) and $\mathbf{KT}_6$(*Router Interference*) in the training set are similar, but there is a data imbalance problem. In the current network, $\mathbf{KT}_9$(*NR Systems Interference*) and $\mathbf{KT}_{10}$(*LTE D-band interference*) interference account for the vast majority, and most of them are compound interference samples. For the small sample data such as $\mathbf{KT}_2$(*China Telecom LTE 800MHz Band Interference*), $\mathbf{KT}_3$(*MMDS signal Interference*), and $\mathbf{KT}_7$(*Pseudo Base Station interference*), most of them are independent interference types. $\mathbf{KT}_8$(*Atmospheric Dust Interference*) is a type of interference with more small-sample compound types.

In Tab. 4, the proportion of $\mathbf{KT}_5$ in the T_B test set increases, and most of them are compound Interference. $\mathbf{KT}_7$, $\mathbf{KT}_1$, $\mathbf{KT}_4$ and other compound interference samples also increased. Therefore, compared with the province A training set, the interference situation in province B is more complicated.

TABLE 3. Statistics of province a training set.

Interference Type	All type samples	Single type sample	Compound Type sample
$\mathbf{KT}_1$	527	522	5
$\mathbf{KT}_2$	29	29	0
$\mathbf{KT}_3$	79	79	0
$\mathbf{KT}_4$	1838	1382	456
$\mathbf{KT}_5$	1771	1684	87
$\mathbf{KT}_6$	534	103	103
$\mathbf{KT}_7$	51	46	5
$\mathbf{KT}_8$	67	35	32
$\mathbf{KT}_9$	5454	1177	4277
$\mathbf{KT}_{10}$	4923	1079	3844

TABLE 4. Statistics of province b test set.

Interference Type	All type samples	Single type sample	Compound Type sample
$\mathbf{KT}_1$	5	1	4
$\mathbf{KT}_2$	1	1	0
$\mathbf{KT}_3$	0	0	0
$\mathbf{KT}_4$	121	58	63
$\mathbf{KT}_5$	441	170	271
$\mathbf{KT}_6$	52	5	47
$\mathbf{KT}_7$	2	1	1
$\mathbf{KT}_8$	44	15	29
$\mathbf{KT}_9$	767	203	564
$\mathbf{KT}_{10}$	180	7	173

B. EXPERIMENT ANALYSIS

We conduct algorithm performance verification based on the above three datasets. According to the requirements of compound interference identification in the current network, new performance metrics, ALL and ANY accuracy, are proposed. Because the interference source is divided into in-system interference and out-of-system interference. If the signal intensity of the in-system interference is weak, engineers may not be able to fix the problem right away. Therefore, the

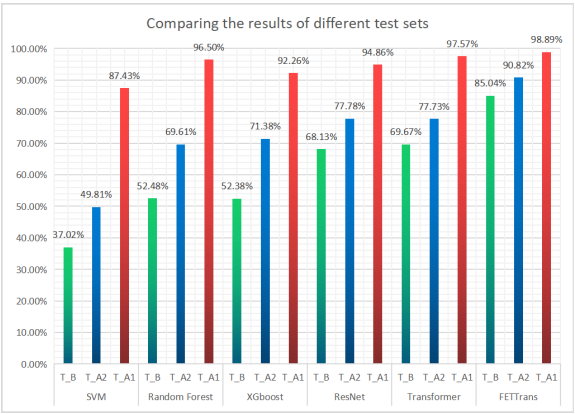


FIGURE 8. Comparison results of algorithms on different test sets.

review results sometimes focus on the correct classification of a subset of labels. ALL accuracy indicates the number of classifications and the correct type of classification, and ANY accuracy rate indicates that the output is a subset of the correct labels.

In addition, the algorithm performance metrics also compare mAP (mean average precision), and metrics for evaluating multi-label tasks such as Hamming loss, Coverage, and Ranking loss [13].

The hyperparameters in the FETTrans network are: (a) block_num of TrEnc-Block and TrDec-Block is set to 4. (b) MultiHead is set to 4. (c) dropout is set to 0.5 and lr is set to 0.002.

1) ALGORITHM COMPARISON RESULTS

The mAP metric can objectively represent the recognition accuracy of the algorithm in each category. Fig. 8 shows the mAP results of different algorithms on different test sets. In this paper, we compare the performance of SVM, Random Forest, XGBoost, ResNet, Transformer, and FETTrans algorithms. Some of them are the baseline algorithms. And we have done feature engineering work on the data and fine-tuned the algorithm parameters. For example, we shift the data with a mean of zero, because the mean of the original data is around -110 dBm. And on this basis, the gradient information of the data is added. According to the method verified by the engineer through the visual image, we also tried to scale the visual image pixel to 224*224 size and then tested it based on the image classification algorithm ResNet-34 network. We did not use a deeper network because the performance decreased as the number of layers increased. The Transformer algorithm results are based on the naive Transformer network. The model takes the output of the FEt module as input and is directly trained as a multi-classifier.

The abscissa in Fig. 8 represents the validation of different algorithms on three test sets. To make the comparison algorithm performance more regular, we show them in the order of test set T_B (green bar), T_A2 (blue bar), T_A1 (red bar). For example, the green bars represent mAP results of different

TABLE 5. Algorithm accuracy comparison.

	Test set	ALL	ANY	Average mAP
SVM	T_A1	80.33%	87.22%	58.09%
	T_A2	54.52%	63.86%	
	T_B	29.72%	53.53%	
Random Forest	T_A1	85.61%	90.66%	72.86%
	T_A2	56.97%	65.54%	
	T_B	37.66%	64.28%	
XGBoost	T_A1	86.65%	91.60%	72.01%
	T_A2	60.49%	70.14%	
	T_B	40.76%	68.54%	
ResNet	T_A1	90.2%	95.8%	80.26%
	T_A2	61.6%	89.9%	
	T_B	54.1%	91.3%	
Transformer	T_A1	92.78%	96.32%	81.66%
	T_A2	69.98%	85.76%	
	T_B	64.70%	85.87%	
FETTrans	T_A1	93.72%	96.04%	91.58%
	T_A2	80.72%	90.59%	
	T_B	73.57%	90.22%	

TABLE 6. Comparison of metrics for multi-label tasks.

	Test set	Hamming Loss	Coverage	Ranking Loss
SVM	T_A1	0.026	1.833	0.092
	T_A2	0.071	1.997	0.222
	T_B	0.122	2.856	0.357
Random Forest	T_A1	0.022	1.532	0.009
	T_A2	0.084	1.899	0.056
	T_B	0.100	2.263	0.064
XGBoost	T_A1	0.019	1.551	0.007
	T_A2	0.066	1.835	0.039
	T_B	0.091	2.430	0.063
ResNet	T_A1	0.015	1.515	0.005
	T_A2	0.050	1.744	0.032
	T_B	0.067	2.156	0.044
Transformer	T_A1	0.015	1.519	0.005
	T_A2	0.051	1.711	0.029
	T_B	0.067	2.080	0.038
FETTrans	T_A1	0.009	1.476	0.003
	T_A2	0.032	1.687	0.029
	T_B	0.042	1.957	0.031

algorithms on the T_B test set. From the results, it can be concluded that all algorithms have acceptable recognition performance on the T_A1 test set. But from the T_A2 and T_B test set validation, we found that all algorithms have different degrees of performance degradation. And for the T_B, the recognition performance becomes worse.

The following conclusions can be drawn from the results in Fig. 8. First, although the algorithm capabilities are

different, the interference type frequency band characteristics can be learned. Second, from the results of the T_A2 and T_B test sets, we can see that the algorithm is overfitting. Third, from the comparison results of T_A1 and T_A2, it is found that there are a lot of uncertain factors in the current network, so it is very important to learn critical frequency band features in the limited training set. And for this task, the ensemble algorithm is relatively more advantageous because its generalization ability is also more obvious.

Finally, the recognition performance of the FETTrans network for T_A2 and T_B test sets degrades, but the average mAP of the three test sets can reach 91.58%, and the mAP for T_B is more than 15% higher than other algorithms. Therefore, FETTrans has higher classification performance and more obvious generalization performance.

Tab. 5 compares the ALL and ANY accuracy of the algorithms and the average mAP values of different test sets. The classification results of ALL and ANY are based on the threshold of 0.5, and if the classification result is empty, the maximum classification probability result will be output. The Average mAP is obtained by averaging the mAP values of different test sets according to Fig. 8.

From the results in Tab. 5, we can conclude that there is a serious overfitting phenomenon for the SVM and ELM algorithms. ResNet is more inclined to output a single result, and has a greater confidence in a single interference in a compound interference.

Tab. 6 is similar to Tab. 5, and the comparison results are the metrics used to evaluate multi-label task. The threshold of hammering loss is also set to 0.5.

2) RECOGNITION PERFORMANCE FOR DIFFERENT TYPES OF INTERFERENCE

We analyze the performance of each type of interference classification algorithm through Tab. 7. Among them, XGBoost is the experimental result of the control group. We found that for easily confused $\mathbf{KT}_4$ and $\mathbf{KT}_6$. Because $\mathbf{KT}_6$ has fewer samples than $\mathbf{KT}_4$, its AP value is also lower. For only one $\mathbf{KT}_2$ type sample in the T_B test set, it will have a greater impact on mAP. In addition, for the small samples of $\mathbf{KT}_2$ and $\mathbf{KT}_7$, their signal overlap is likely to produce a frequency band waveform similar to that of $\mathbf{KT}_{10}$. This resulted in lower AP values for $\mathbf{KT}_2$ and $\mathbf{KT}_7$. The full-band interference $\mathbf{KT}_1$ interference, which is a small sample, is more difficult to classify because it is easily confused with $\mathbf{KT}_5$. From three comparative experiments, the bidirectional dynamic adaptive approach of the FETTrans network significantly improves these situations.

V. CONCLUSION

According to the current network data in different periods and different regions, this paper analyzes the classification of compound uplink interference. Due to the diverse distribution of the same interference type and the superposition of interference signals, the FETTrans network is proposed. The network structure based on the Transformer algorithm enables FETTrans to transform the compound interference identification problem into a natural language processing task. By constructing the Key-Token of the output subspace and the input feature Data-Token for feature adaptation, the bidirectional dynamic training process of the network is realized. This adaptation process solves the decoupling of the composite interfering signal. Therefore, the FETTrans network can effectively solve the multi-label classification task of compound interference. By introducing Key-Token, the algorithm can effectively improve recognition accuracy and generalization performance.

The FETTrans network implements an end-to-end algorithm and can complete training and inference by changing the output type. In the comparison of the results of multiple algorithms, the FETTrans algorithm has a 15% improvement in the accuracy of the classification algorithm and excellent generalization ability. At present, FETTrans focuses more attention on the process of feature adaptation, and the future algorithm network can further focus on the distinction between data features. The accurate identification of compound interference sources is the cornerstone of the map coordinate positioning of the interference sources, and it is also the next work plan.

APPENDIX.

This section introduces 10 common types of interference at this stage. From the data point of view, we divide it into

TABLE 7. Average precision for each interference type.

	Test set	$\mathbf{KT}_1(\%)$	$\mathbf{KT}_2(\%)$	$\mathbf{KT}_3(\%)$	$\mathbf{KT}_4(\%)$	$\mathbf{KT}_5(\%)$	$\mathbf{KT}_6(\%)$	$\mathbf{KT}_7(\%)$	$\mathbf{KT}_8(\%)$	$\mathbf{KT}_9(\%)$	$\mathbf{KT}_{10}(\%)$
XGBoost	T_A1	94.44	90.91	100	94.3	93.01	83.43	92.31	84.62	96.05	93.48
	T_A2	59.33	61.47	-	70.05	84.11	39.62	-	-	97.09	88.02
	T_B	33.14	7.42	-	83.69	92.52	48.6	31.44	9.02	93.77	71.81
Transformer	T_A1	96.76	100	100	98.32	99	93.34	98.97	93.78	98.54	97.01
	T_A2	76.74	76.02	-	64.81	83.96	56.24	-	-	97.75	88.6
	T_B	57.88	1.41	-	86.04	96.47	63.41	36.19	100	94.15	91.56
FETTrans	T_A1	98.01	100	100	99.11	99.06	98.14	100	95.45	99.7	99.43
	T_A2	75.91	100	-	92.26	94.36	83.33	-	-	97.6	92.34
	T_B	65.18	85.04	-	91.88	95.42	77.89	64.03	100	94.4	91.49

TABLE 8. Full-band interference.

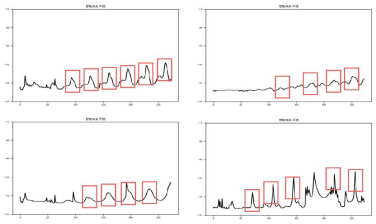
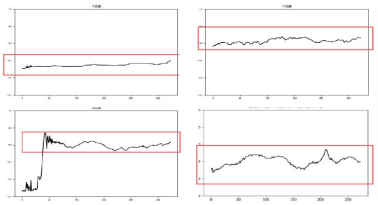
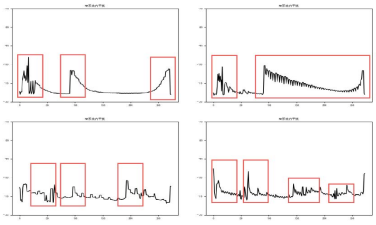
	Interference source	Principle of interference generation	Interference visualization
full-band interference	**Intelligent Street Lamp Interference** •Interference source: Intelligent Street Lamp •Common places:streets, highways **Interference feature** In most cases, the waveform is generally high on the right and low on the left, and fluctuates peri-odically. Sometimes also manifest-ed in part of the frequency band.	The wireless control module included in the intelligent st-reet can continuously and illegally transmit a signal with a center frequency of 2620MHz. And the bandwidth is about 150M, which is used to detect the distance of the human body to automatically adjust the brightness. Such interference is widespread in some regions.	
	Jammer Interference • Interference source:Jammer • Common places:prison, school **Interference feature** From the perspective of frequency domain characteristics, it generally shows an overall noise floor increase or a large bandwidth noise floor increase.	Around the prison/school, jammers are often turned on to interfere with the entire wireless signal.	
	NR Systems Interference •Interference source: Sites with high uplink load in the system •Common places:Base station overlapping coverage area **Interference feature** The data is distributed in multiple frequencies with a stair-like wave-like slow-down feature	With the large increase of 5G users, the traffic channel l-oad is significantly increased. In high traffic areas, the tr-affic load or the unreasonable setting of the PUSCH fun-ction parameters of the cell will cause the PUSCH chan-nel interference to increase significantly. Different manufacturers enable the feature of randomiza-tion of interference, and allocate resources from differe-nt RB positions of PUSCH, which leads to the diversity of interference in the NR system.	

TABLE 9. Narrowband of fixed-band interference.

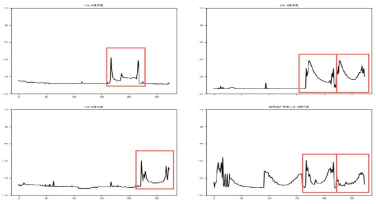
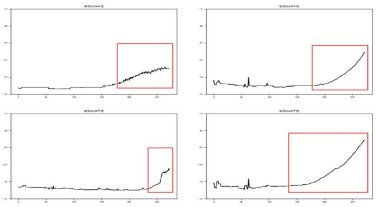
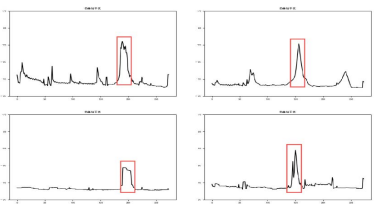
	Interference source	Principle of interference generation	Interference visualization
narrowband of fixed-band interference	**LTE D-band interference Interference** •Interference source: The uplink terminal of the LTE cell •Common places:Areas where LTE frequency clearing has not been completed **Interference feature** The noise floor of 163-273 PRB is obviously raised, and there are two 20M bandwidth/single 20M interference in the right D1/D2 frequency band	Due to the problem of spectrum division, the original LT-E D-band is 2575-2615MHz, and the LTE cell with 40M bandwidth needs to be shifted to 2615-2655MHz. Howe-ver, the current traffic volume of the LTE D-band is rela-tively high, and some LTE D-band cells have not been completely de-frequencyed, causing interference to the 5G system. Depending on the LTE de-frequency situati-on, there may be single-carrier and dual-carrier interfer-ence.	
	China Telecom LTE 800MHz Band Interference • Interference source: China Telecom company 800M third-order intermodulation •Common places: Guangxi Province **Interference feature** The frequency band of 2515-2615Hz shows the charact-eristics of spurious interference in the right band.	800M downlink spectrum of China Telecom is 869MHz -880MHz, and the frequency range of the third harmonic is 2607MHz-2640MHz. If the intermodulation interfer-ence suppression is not good, some signals will fall into the 2515MHz-2615MHz frequency band, showing the characteristics of spurious interference in the right part of band.	
	Pseudo Base Station Interference •Interference source: pseudo base station • Common places:near the police station **Interference feature** Waveforms with narrow frequency band characteristics and fixed positions	The LTE D-band pseudo base station uses the 2.6GHz frequency band, which will also interfere with the 5G system. The bandwidth of the pseudo base station can support 1.4/3/5/10MHz. The common bandwidth of the current network is 5MHz.	

TABLE 9. *(Continued)* **Narrowband of fixed-band interference.**

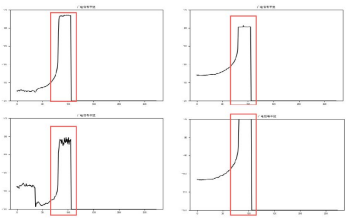
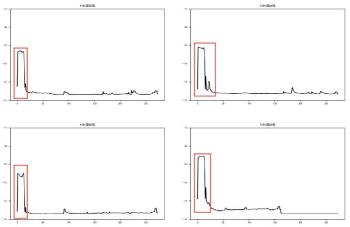
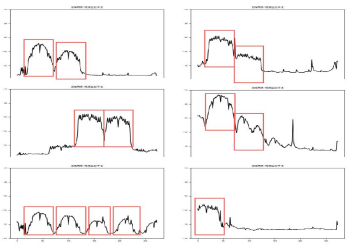
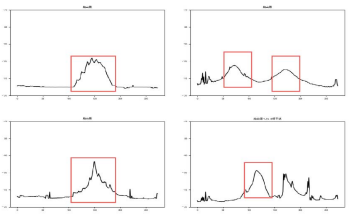
narrowband of fixed-band interference	**MMDS signal Interference** •Interference source: National Radio and Television signal towern •Common places:Some city radio and television companies have not withdrawn the frequency **Interference feature** As shown on the right figures	Multi-channel Microwave Distribution System (MMDS), which uses microwave signals to transmit radio and televi -sion signals. The radio and television signals in the 2.6 GHz band have been de-frequencyed in a large area, and the slow progress of a few cities still has a small impact. The new spectrum of 700MHz has a large amount of radio and television signal interference. From the frequency domain characteristics, the interference characteristics are mainly square waves with a width of about 8MHz.	
	Atmosphere Duct Interference •Interference source: Atmosphere under certain conditions of temperature, climate, topography, etc. • Common places:appear randomly **Interference feature** The first 12 RBs suffer from higher interference	Atmospheric duct interference is strictly a kind of ultra -distance inter-ference. Under certain conditions of temperature, climate, terrain, etc. in the atmosphere, the downlink signal transmission of the disturbing base station exceeds the guard time slot of the uplink signal of the disturbed base station due to the reduction of the propation loss of the wireless signal. The interference signature in principle presents a wide bandwidth rise in the frequency domain. However, due to the current low 5G traffic in some areas and low PDSCH traffic channel load, more common channel interference features, such as SSB signals; In the time domain, the ultra-far interfer-ence exhibits a ramp-like slow decline. In addition, because the LTE D-band is not completely defrequencyed, atmospheric waveguides are often accompanied by 4G signals.	

TABLE 10. **Narrowband of flexible interference.**

	Interference source	Principle of interference generation	Interference visualization
narrowband of flexible interference	**Wireless Bridge / Video Surveillance Interference** **Interference** •Interference source: Video surveillance wireless bridge, wireless backhaul and other equipment • Common places:Buildings, communities, such as elevators, parking lots, etc.s **Interference feature** It may exist in the entire 2.6GHz 100M range, showing an interference increase of about 50RB, which may be single or multiple	Video surveillance interference generally refers to the illegal occupation of the 2.6G frequency band by the video wireless backhaul part.	
	Router Interference •Interference source: Illegal installation of routers at indoor sites •Common places:Indoor hidden spot **Interference feature** From the perspective of frequency domain interference characteristics, full-band interference with lower width and upper tip is likely to occur in the entire 100M frequency band of 2.6GHz	When a wireless router performs poorly, it often causes signal leakage in the 2.6GHz band, which interferes with the 5G signal.	

full-band interference and narrowband interference, and narrowband interference is further divided into fixed-band interference and flexible interference. The full-band interference means that the interference source frequency band affects the entire frequency band of the communication frequency band (2515MHz-2675MHz), and the narrowband means that the interference source will affect part of the communication frequency band. According to the center position of the frequency band, it is subdivided into fixed-band interference and flexible interference.

Tab. 8 represents the full-band interference, the second column is the introduction of the interference source, the third column is the interference principle, and the fourth column is the interference feature circled by the red frame. By analogy, Tab. 9 represents narrowband of fixed-band interference, and Tab. 10 represents narrowband of flexible interference.

REFERENCES

[1] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. New York, NY, USA: Academic, 2020.

[2] M. Schmidt, D. Block, and U. Meier, "Wireless interference identification with convolutional neural networks," in *Proc. IEEE 15th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2017, pp. 180–185.

[3] S. Grunau, D. Block, and U. Meier, "Multi-label wireless interference identification with convolutional neural networks," 2018, *arXiv:1804.04395*.

[4] J. Webber, K. Yano, N. Suga, Y. Hou, E. Nii, T. Higashimori, A. Mehbodniya, and Y. Suzuki, "WLAN interference identification using a convolutional neural network for factory environments," *J. Commun.*, vol. 16, no. 7, pp. 276–283, 2021.

[5] S. Grimaldi, A. Mahmood, and M. Gidlund, "Real-time interference identification via supervised learning: Embedding coexistence awareness in IoT devices," *IEEE Access*, vol. 7, pp. 835–850, 2019.

[6] J. Shen, S. Li, F. Jia, H. Zuo, and J. Ma, "A deep multi-label learning framework for the intelligent fault diagnosis of machines," *IEEE Access*, vol. 8, pp. 113557–113566, 2020.

[7] A. Qin, Q. Hu, Y. Lv, and Q. Zhang, "Concurrent fault diagnosis based on Bayesian discriminating analysis and time series analysis with dimensionless parameters," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2254–2265, Mar. 2019.

[8] F. Li, X. Ma, and Y. Wang, "A multi-label method of state partition and fault diagnosis based on binary relevance algorithm," in *Proc. IEEE 9th Data Driven Control Learn. Syst. Conf. (DDCLS)*, Nov. 2020, pp. 567–572.

[9] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, and H. Cho, "XGBoost: Extreme gradient boosting," *R Package Version* vol. 1, no. 4, pp. 1–4, Aug. 2015.

[10] Q. Hu, A. Qin, Q. Zhang, J. He, and G. Sun, "Fault diagnosis based on weighted extreme learning machine with wavelet packet decomposition and KPCA," *IEEE Sensors J.*, vol. 18, no. 20, pp. 8472–8483, Oct. 2018.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

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

[13] X. Zheng, P. Li, Z. Chu, and X. Hu, "A survey on multi-label data stream classification," *IEEE Access*, vol. 8, pp. 1249–1275, 2020.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[15] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov, "Transformer-XL: Attentive language models beyond a fixed-length context," 2019, *arXiv:1901.02860*.

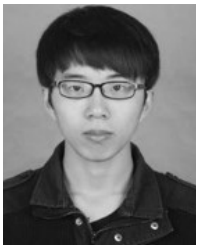
[16] M. Boroumand, M. Chen, and J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 5, pp. 1181–1193, May 2018.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[18] I. Beltagy, M. E. Peters, and A. Cohan, "Longformer: The long-document transformer," 2020, *arXiv:2004.05150*.

[19] N. Park and S. Kim, "How do vision transformers work?" 2022, *arXiv:2202.06709*.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, vol. 1, Jun. 2015, p. 9.



BORAN LI was born in 1993. He received the master's degree from Northeastern University, Shenyang, China, in 2020. He is currently a Senior System Developer with the Wireless and Terminal Technology Research Institute, China Mobile Research Institute. His research interests include telecom wireless network optimization solution and artificial intelligence algorithm.

LEI ZHANG was born in 1984. He received the bachelor's degree from Xidian University, Xi'an, China, in 2010. He is currently a Senior System Developer with the Wireless and Terminal Technology Research Institute, China Mobile Research Institute. He has seven years of experience in wireless network maintenance and optimization. His research interests include telecom wireless network optimization solution and artificial intelligence algorithm.

JINGWEI DAI was born in 1985. She received the bachelor's and master's degrees from Beijing Jiaotong University, Beijing, China, in 2009 and 2012, respectively. She is currently a Senior System Developer with the Wireless and Terminal Technology Research Institute, China Mobile Research Institute. Her research interests include telecom wireless network optimization solution and artificial intelligence algorithm.

JIAN MA was born in 1984. He is currently the Technical Manager of the Wireless and Terminal Technology Research Institute, China Mobile Research Institute. His main research interests include intelligent optimization of wireless networks, wireless network management functions, and product development.

ZHIMIN ZHAO is currently a Senior Engineer with China Mobile Guangxi Company, a junior expert in network operation, engaged in wireless network optimization for 18 years. His main research interest includes 4G/5G network maintenance and optimization.