A New Attention-based Neural Network for The Identification of Non-line-of-sight Signals in Data from Global Navigation Satellite Systems

Y. F. Zhao, Z. Q. Dai, F. Li, X. W. Zhu, C. X. Ran

***Abstract*—Modern satellite positioning primarily uses global navigation satellite system (GNSS) signals collected by multiple satellites and obtains specific location information through calculations. However, the nonlinear propagation of the signal between the satellite and receiver caused by reflection or scattering due to buildings forms a non-line-of-sight (NLOS) signal, which significantly degrades positioning accuracy. Therefore, distinguishing and eliminating NLOS signals is the best way to improve the accuracy of modern satellite navigation. Initially, algorithms and physical models were established to eliminate NLOS signals. Then, some machine learning models were developed that appeared to be able to better solve the problem, such as the gradient boosting decision tree, support vector machine, convolutional neural network (CNN), and long short term memory (LSTM) network. However, adapting the same model to different environments and the fine extraction of the original features of the signal are challenges that remain unresolved. Therefore, this paper proposes a neural network based on the attention mechanism, called the environmental transformer (ET), that can extract both the satellite visibility features of the signal and the environmental features around the GNSS signal receiver. In the binary classification task of distinguishing between NLOS and LOS signals, an ET can reach an accuracy of 87.45%, which is higher than that of previously mentioned models. The results also show that the adaptability of an ET to different environments has greatly improved. This paper also explains the working principle of an ET through attention visualization.**

***Index Terms*—Attention mechanism, deep learning, GNSS, NLOS, signal visibility classification**

I. Introduction

As society continues to develop, the environment in which users use positioning has become more complex. In the first Fresnel zone of a global navigation satellite system (GNSS) signal receiver, an object blocks the radio waves, causing direct waves to be reflected with a phase difference. This signal is called the non-line-of-sight (NLOS) signal, and its formation process is shown in Fig. 1. At this time, the wireless signal can only reach the receiving end through reflection, scattering, and diffraction. This multipath effect causes a series of problems, such as asynchronous delay, signal attenuation, polarization change, and link instability. When the antenna height of the GNSS signal receiver is low and the propagation distance is long, the path difference between the direct wave and the reflected wave will severely reduce the satellite positioning accuracy.

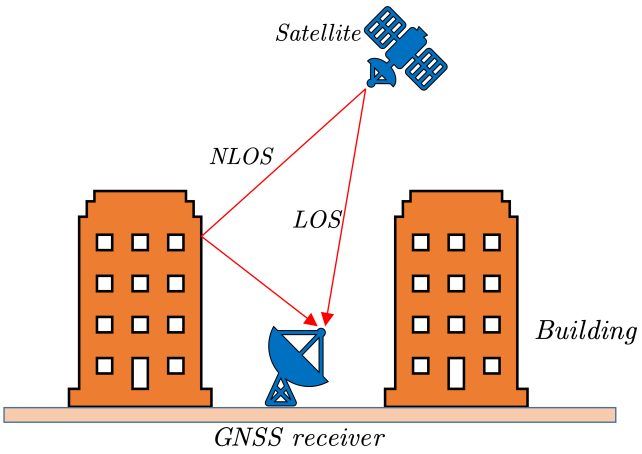


Fig. 1. Formation process of an NLOS signal.

Currently, the main method used to reduce the error caused by the NLOS signal on positioning is to identify the NLOS signal from the GNSS signal receiver and eliminate it. In [1], one of the most direct methods mentioned is to model the environment or develop new algorithms to identify the NLOS and line-of-sight (LOS) signals. In [2], the common algorithms mentioned include residual tests, LS algorithms, identify and discard techniques, and robust estimators. However, as the surrounding environment becomes more complicated, physical modeling and algorithms become extremely difficult. In [3], another proposed method is to use other indirect quality information (such as the carrier-to-noise ratio) to detect those measurement values that contain multipath errors. However, with the continuous changes in the external environment, in most cases, many multipath or NLOS errors do not follow the expected behavior.

Even though traditional mathematics and modeling cannot solve the problem of eliminating NLOS signals, recently,

Manuscript received June 15, 2023; revised September 12, 2024.
This work was supported in part by the Sun Yat-Sen University.

Y. F. Zhao is an undergraduate student at School of Electronics and Communication Engineering, Sun Yat-Sen University, Shenzhen 528406, China (e-mail: zhaoyf37@mail2.sysu.edu.cn).

Z. Q. Dai is an assistant professor of School of Electronics and Communication Engineering, Sun Yat-Sen University, Shenzhen 528406, China. (corresponding author, e-mail: daizhiqiang@mail.sysu.edu.cn).

F. Li is a postgraduate at School of Electronics and Communication Engineering of Sun Yat-Sen University, Shenzhen 528406, China. (e-mail: lifang63@mail2.sysu.edu.cn).

X. W. Zhu is a professor at School of Electronics and Communication Engineering, Sun Yat-Sen University, Shenzhen 528406, China. (e-mail: zhuxw666@mail.sysu.edu.cn).

C. X. Ran is a professor at School of Electronics and Communication Engineering, Sun Yat-Sen University, Shenzhen 528406, China. (e-mail: ranchx@mail.sysu.edu.cn).

popular machine learning algorithms have provided a new solution that can model the problem even without algorithmically simulating the complex principles, allowing machines to replace humans in classifying the visibility of signals in terms of tuning and learning parameters. In what follows, we first introduce some commonly used machine learning methods.In [4–5], research based on the gradient boosting decision tree showed a relatively high accuracy rate that could be achieved for signal reception classification in contrast with the distance-weighted k-nearest neighbor and adaptive network-based fuzzy inference systems. Moreover, research based on support vector machines has achieved good results[6–9]. However, in the face of such a complicated black-box problem, the tuning of machine-learning models remains a big problem. Recently, deep learning technology has been developed rapidly, and it can be seen as a smarter mechanism than machine learning, as it can learn more advanced functions from the given data by continuously deepening the neural network. Next, we introduce the application of common deep-learning neural networks in signal visibility classification. In [10], the researchers proposed using the recurrent neural network model, which takes a series of channel state information to identify the corresponding channel condition. This paper uses the CNN-LSTM module, into which CIR data were directly input. In [11], a CNN was employed for exploring and extracting features automatically, and then the CNN outputs were fed into the LSTM network for classification. In [12], researchers also used the combination of a fully convolutional neural network (FCNN) and LSTM to extract signal visibility features and signal timing features simultaneously. Transfer learning can also be applied to signal visibility classification. Another paper [13] proposed a transfer-learning-based NLOS identification method for classifying the NLOS.

Although the recognition rate of NLOS signals using many deep learning techniques is already excellent, some problems remain that have not been resolved, such as identifying the relationship between the different dimensions of the feature vector of the input neural network or learning how to make the same neural network show strong adaptability in different environments. Therefore, to better solve the above problems, this paper proposes a new deep learning network based on the attention mechanism in [14], which can not only extract signal visibility features in complex environments but also enhance the adaptability of the network to different environments. This network is inspired by the transformer, and thus called the environmental transformer (ET).

The contributions of this paper are as follows. First, this paper proposes a new neural network based on the attention mechanism. Second, to improve the adaptability to the environment, this paper uses the environment feature extraction network. Finally, this paper establishes the relationship between different dimensions in a sample vector through the attention weight visualization.

The rest of this article is organized as follows. In Section 2, this paper describes the proposed ET neural network in detail and explains the reasons for the setting of each part of the neural network. In Section 3, this paper extracts the feature vectors of five dimensions through calculations for the network to perform feature extraction. In Section 4, the binary classification results from multiple perspectives are analyzed. In Section 5, the effect of different dimensional feature vectors through attention weight visualization is analyzed. We explain the reasons for choosing the attention mechanism to construct and verify the improvement of the environmental feature extraction network for environmental adaptability. Finally, Section 6 summarizes the whole article and puts forward an application prospect.

II. System Architecture

The proposed ET neural network architecture for GNSS satellite visibility is shown in Fig. 2. To initially extract the characteristics of the NLOS and LOS signals, we preprocess the original GNSS signal into a five-dimensional feature vector consisting of pseudo-range measurement, Doppler measurement, the azimuth angle, the elevation angle, and the carrier-to-noise density ratio. Then, we call the number of feature vectors input to the neural network each time the batch size, and input all the feature vectors in a batch into the environmental block to extract the environmental information around the GNSS receiver. At the same time, we input each feature vector into the encoder block, extract the satellite visibility information from it through the attention mechanism, and finally combine the two to obtain a GNSS satellite visibility prediction that is adaptable to the environment. As shown in Fig. 2, the ET deep learning neural network draws on the ideas of the transformer model in [14] and the encoder-decoder model[15–16] using the adapted transformer's encoder layer, the self-made environmental layer, and the decoder layer to obtain the GNSS satellite visibility classification results. Next, this paper introduces the specific details of each part of the ET neural network and the reason for the setting of each part.

A. Attention mechanism

Since this is the first time the attention mechanism has been introduced in satellite visibility classification, we first introduce the attention mechanism [16–18].

a. The contrast between attention and non-attention

Non-attention: For example, in the traditional FCNN layer, each input has its corresponding output, and the finally learned network is equal to each input. This undifferentiated feature extraction is called non-attention feature extraction.

Attention mechanism: This involves paying attention to queries. For any given query, the attention mechanism uses attention for pooling and biases the selection toward sensory input. In the context of the attention mechanism, these sensory inputs are called key-value pairs. Each key corresponds to a value, and this correspondence is non-attentive. Different attention mechanisms can be designed to allow different queries to interact with key-value pairs that are attentive. In this way, queries can extract more subtle features from sensory input. This feature extraction mechanism is called attention.

b. Attention mechanism

This paragraph explains the attention mechanism in Fig. 3. First, enter the ith query and m key-value pairs. Next, use the α function to represent the importance of the key to the query, which expresses the attention weight, as shown in (2) and (3). Then, after passing through the soft-max layer, the attention weight is multiplied by the value to obtain the probability distribution of the value corresponding to each key. Finally, the attention output is obtained after adding m results. In [16], the whole process is called attention pooling, as shown in (1)

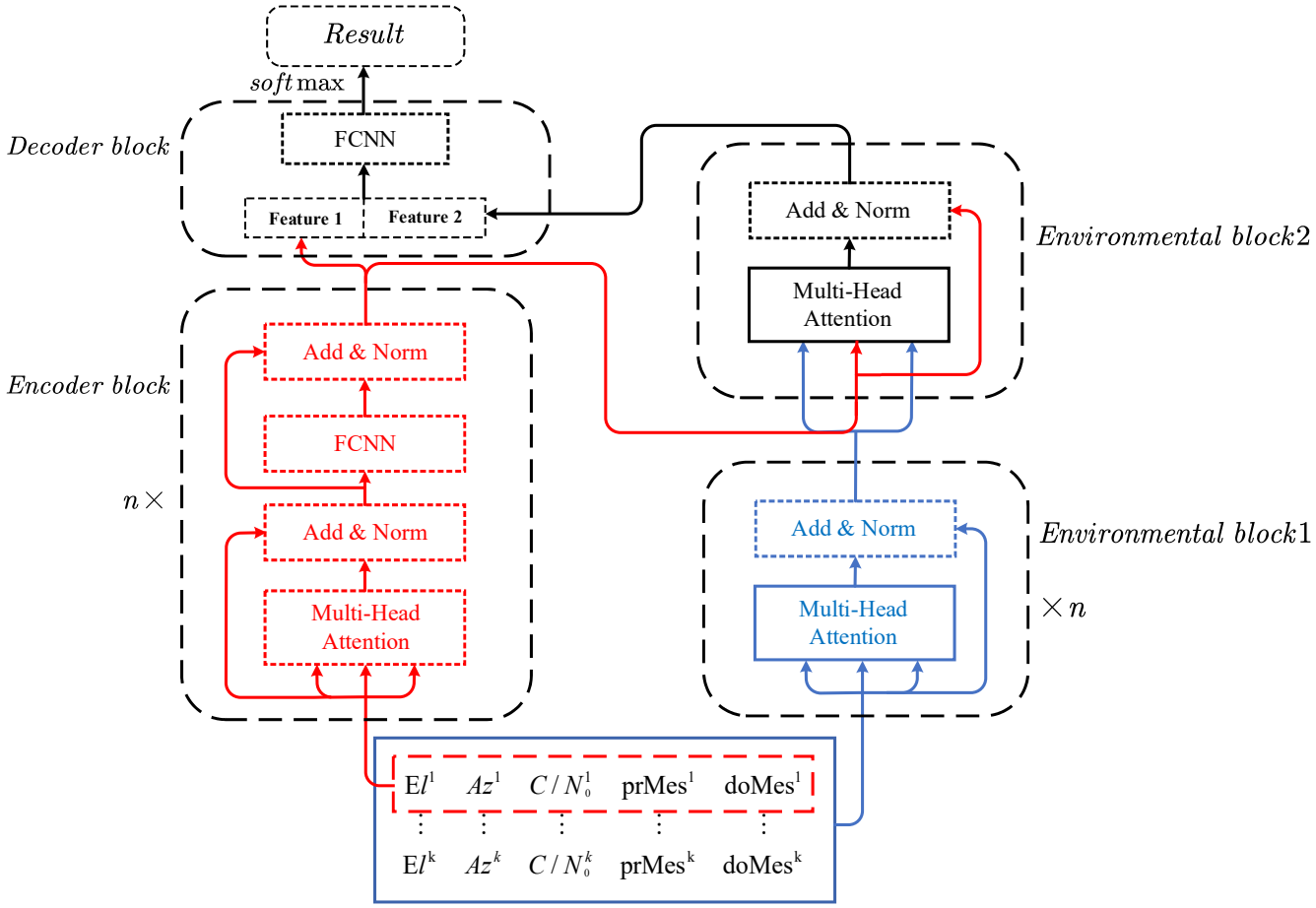


Fig. 2. The basic architecture of the ET neural network.

and (4). The attention mechanism used here is additive attention.

$$soft\max(x_k) = \frac{e^{x_i}}{\sum_{i=1}^{m} e^{x_k}} \ (i \in [1, m]) \tag{1}$$

$$\tanh(x) = \frac{\sinh x}{\cosh x} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2}$$

$$\alpha(q,k) = w_v^T \tanh(W_q * q + W_k * k) \tag{3}$$

$$f(q, (k_i, v_i)) = \sum_{i=1}^{m} soft\max(\alpha(a, k_i)) v_i \tag{4}$$

($W_q \in \mathbb{R}^{h \times q}; W_k \in \mathbb{R}^{h \times k}; w \in \mathbb{R}^{h}$ are learnable parameters)

In short, the output of attention pooling is the result of queries having a different weight for each key-value pair after the attention evaluation has been performed on all key-value pairs. Attention pooling includes the concept of active search features.

B. Encoder block

As shown in Fig. 2, the encoder block contains four parts. First, to extract the visibility characteristics of satellite signals, each sample is input independently into the multi-head attention. Then, the add&norm block is used to prevent gradient explosion or overfitting problems in the model. Next, through the FCNN block, a linear factor is added to the model and the output size is adjusted. Finally, the add&norm block is set again to normalize the data for further training. By continuously optimizing the model, we can set n to 2, which represents two cycles of training carried out. Next, we focus on the multi-head attention mechanism and the add&norm block in the encoder block.

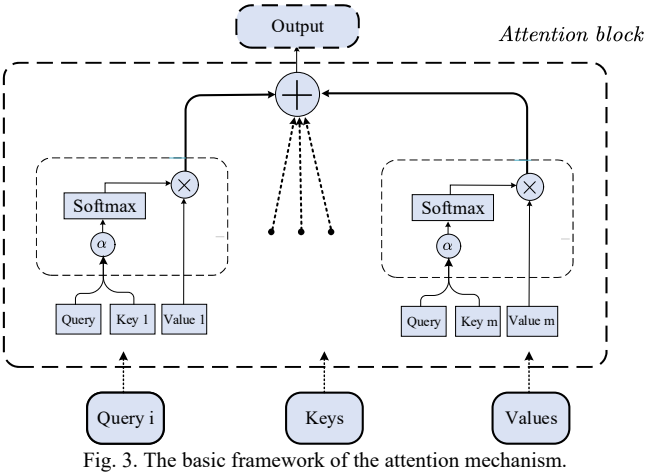


Fig. 3. The basic framework of the attention mechanism.

a. Multi-head attention block

To obtain the satellite signal visibility feature from different angles, the model can learn different behaviors based on the same attention mechanism and then combine the different behaviors as output. As shown in Figure 4, n copies of queries, keys, and values are copied and then input into different attention blocks. Each attention block, shown in Fig. 3, is called an attention head, and "num heads" represent the number of attention blocks. Then, the output of n attention layers is spliced, the dimensionality of the output is adjusted through the FCNN layer, and finally, the output is obtained.

To this end, multiple sets of different linear projections that have been learned independently can be used. Finally, the different outputs of attention pooling are concatenated into a vector through splicing. Overall, a multi-head attention block can capture all aspects of the feature details so that the outputs are more detailed, comprehensive, and representative.

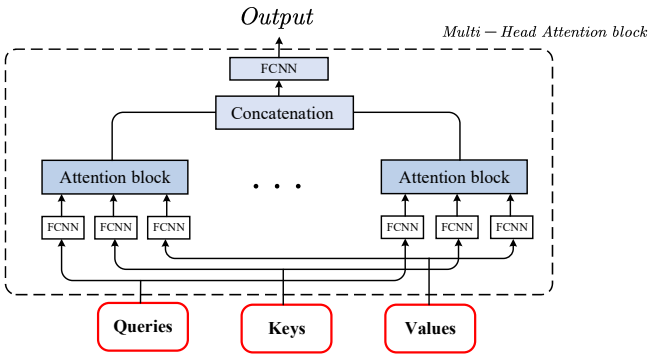


Figure 4. The basic framework of the multi-head attention mechanism.

b. Add&norm block

Inspired by the transformer model, this block consists of two parts, and its purpose is to increase the robustness of the neural network and increase the prevention of over-fitting problems.

● Residual block [19–20]:

Because the neural network based on the attention mechanism is relatively deep, the problems of gradient disappearance and gradient explosion are prone to occur during general training. Jump links can obtain activation from a specific network layer and then quickly provide feedback to the next layer, or even to a deeper layer of the neural network. Therefore, as shown in Fig. 5 and (5), we add residual blocks between the attention blocks.

$$outputs = x + dropout(y) \tag{5}$$

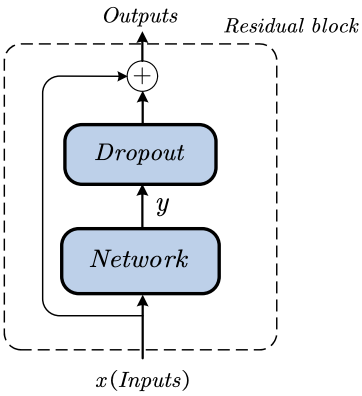


Fig. 5. The basic framework of a residual connection.

The role of the dropout layer is to reduce the complex coadaptation relationship between neurons. Therefore, it is not guaranteed that every two hidden nodes will appear at the same time each time, so the updating of the weights will no longer depend on the common implicit nodes with a fixed relationship. This approach prevents certain features from being effective only under other specific features. From this perspective, a dropout is akin to L1 and L2 regularity, and reducing the weight makes the network more robust to the loss of specific neuron connections.

● Layer normalization [21]:

To overcome the problem of abnormal data scale distribution during the training process, we also adopt the layer normalization method, which can ensure that the mean value in each layer is 1 and that the variance is 0.

C. *Environment blocks*

To increase the adaptability of the network to different environments and to extract the satellite visibility information contained in the environment, we set up two environment blocks. Environment block 1 performs a rough extraction of the surrounding environment to remove some of the noise information from the environment. Environment block 2 makes use of attention to perform the fine extraction of environmental information.

a. Environment block1

As shown in Fig. 2, through the combination of the multi-head attention block and the add&norm block network, all the data in a batch_size are received as a sample to perform the preliminary extraction of data features. At the same time, it is confirmed that the network does not have overfitting problems. All of the data in a batch are input at the same time, meaning that the input contains both NOLS and LOS features. Therefore, the output of this block is not useful for distinguishing NOLS from LOS features, but precisely because the sampling time and sampling location of these similar data are similar, the output of this block contains the environmental information of the sampling location and sampling time. Hence, this block is called the preliminary extraction of environmental information. Through the optimization of the model, we finally set n to 2, which means that using environmental block 1 twice can achieve a better rough extraction of environmental information.

b. Environment block 2

As shown in Fig. 3, the internal structure of environmental block 2 is also composed of a multi-head attention block and add&norm block networks. A feature vector containing all environmental information can be obtained through block 1, using it as the indexed object (keys and values) and using the output of the encoder block as queries to perform the in-depth extraction of environmental features. Therefore, through environmental block 2, each query obtains the corresponding more detailed environmental information. The output of environmental block 2 is the environmental information ultimately used.

D. *Decoder block*

After performing dropout processing on the output of environmental block 2, we splice the output with the output of the encoder block to obtain the NLOS feature information of each signal and its corresponding environmental information. In this block, FCNN adds a linear factor to the network and shrinks its dimension to the size required for the output. Finally, the GNSS satellite visibility classification result is obtained.

E. *Training network*

The goal of deep learning is to adjust the parameters of each layer of learning. These parameters represent the relationship between the existing feature data and the label; thus, when new features are input, the prediction of the label can be achieved. The feature data here refers to the data received from the GNSS receiver, which includes five features. Moreover, the label refers to the result of satellite visibility

classification, which includes the NLOS or LOS signal. In the training process, each batch of feature data is input into the network to obtain the corresponding output. Then, the consistency of the output label and that of the real label are estimated through the loss function. In the ET network, the loss function is defined as the binary cross entropy of satellite visibility prediction. The cross entropy loss value is obtained after 50 epoch training iterations, as shown in Fig. 6.

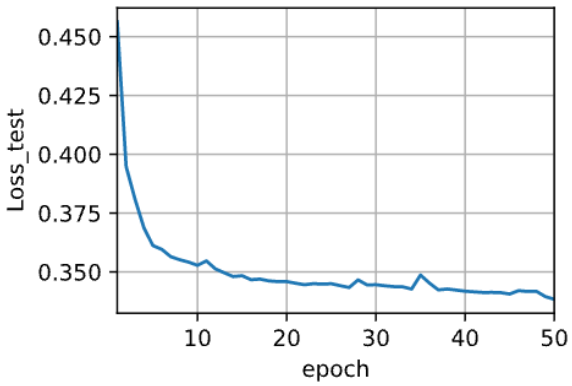
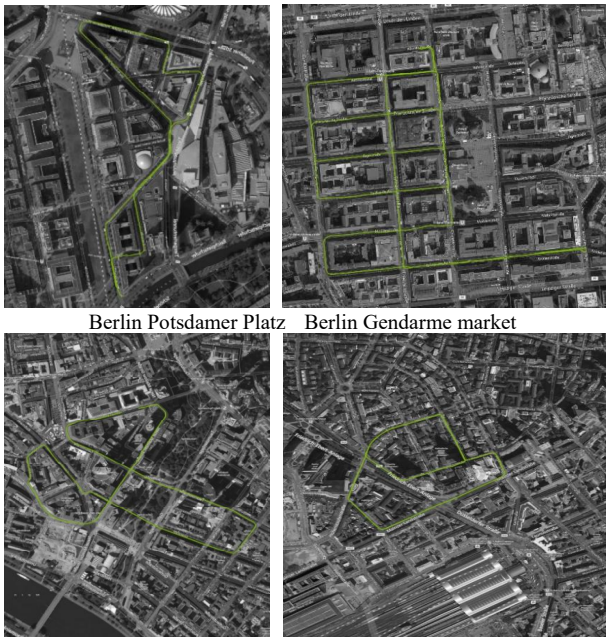


Fig. 6. Cross-entropy loss curve. The abscissa is the number of epochs, and the ordinate is the binary cross-entropy loss.

After calculating the consistency between the output label and the real label, we use back-propagation to estimate the input gradient of each node. In [22], the gradient descent optimization method was used to update each parameter. In this paper, the optimizer chosen is Adam. After applying the trained network, we can predict the corresponding satellite visibility.

III. GNSS Feature Extraction

The dataset used in this experiment is an open-source dataset from the smartLoc project. Fig. 7 is the top view of the data collection site. The specific settings of the experiment collection can be found on the official website of the smartLoc project.



Berlin Potsdamer Platz Berlin Gendarme market

Frankfurt am Main Tower Frankfurt am Main Westend Tower

Fig. 7. Top views of the four data collection sites in Berlin and Franklin.

A deep learning neural network can extract feature information from the raw data received by the GNSS receiver to classify NLOS and LOS signals. However, the original signal contains too much noise, which leads to some network parameters only learning the rules of the noise, or else due to the existence of noise, effective gradient descent cannot be performed. Therefore, to effectively learn the information of all aspects of the signal while ensuring that the characteristics of the signal are not reduced, we summarize the five aspects of the feature dimension: pseudo-range measurement, Doppler measurement, azimuth angle, elevation angle, carrier-to-noise density ratio.

A. Pseudo-range measurement

Pseudo-range can be defined as the difference between the local time the signal is received (i.e., the local clock face time) and the characteristic time value carried by the signal. The measured pseudo-ranges are used to form an equation set with the three-dimensional coordinates of the GNSS receiver antenna phase center and the satellite clock error as unknowns, and the three-dimensional coordinates of the GNSS receiver antenna phase center are then obtained using the least square method. This method can be used for the static or dynamic positioning of GNSS receivers. Therefore, pseudo-range measurement is a feature that the ET neural network should learn.

B. Doppler measurement

The frequency change formed by the Doppler effect is called the Doppler shift, and it is proportional to the relative speed and frequency of vibration. Based on the magnitude of the Doppler frequency shift, the relative velocity of the target's radial movement to the radar can be measured, and according to the time difference between the transmitted pulse and the received time, the distance of the target can be measured. At the same time, the frequency filtering method can be used to detect the Doppler frequency spectrum of the target and filter the spectrum of the interference clutter so that the radar can distinguish the target signal from the strong clutter. Thus, we select Doppler measurement as the feature vector for the ET neural network to learn.

C. Carrier-to-noise ratio

The quality of GNSS measurements is also often evaluated by its corresponding carrier-to-noise ratio, which describes the signal strength of the received signal relative to the noise power density of the receiver front end. Therefore, this paper uses the carrier-to-noise ratio as an input feature to classify the visibility of satellites.

D. Azimuth angle

The azimuth angle of a satellite is a geometric parameter describing its position. Considering the spatial correlation between different satellites, it is only indirectly related to the GNSS measurement quality. This relationship is difficult to express using traditional models, but it may be feasible to use deep learning methods, which are conducive to extracting abstract representations. Hence, we choose the azimuth angle as a feature for the ET neural network to learn. The calculation formula of the azimuth angle is given below.

$$Az = \arctan\left(\frac{\hat{e}_{sv}}{\hat{n}_{sv}}\right) \tag{6}$$

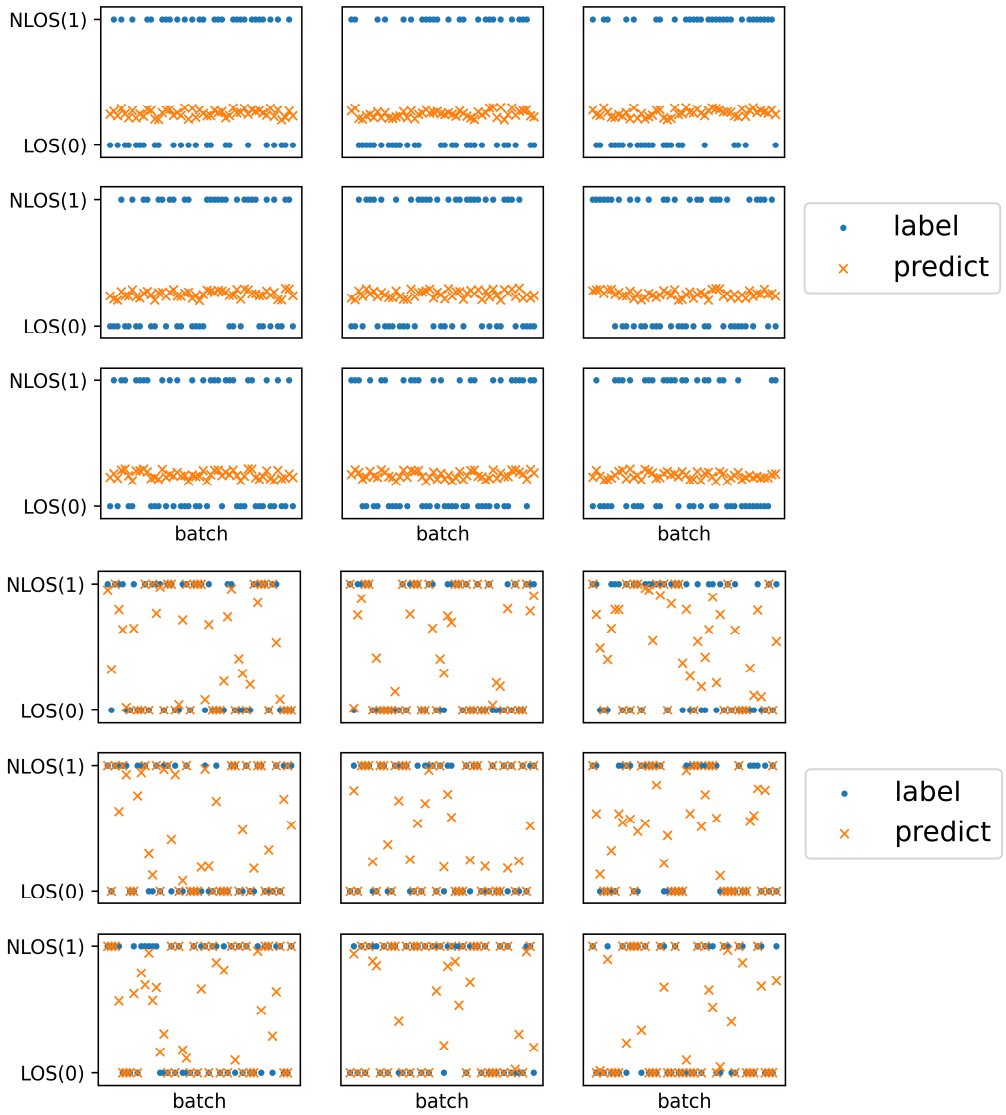


Fig. 8. Process of parameter learning
from the beginning (random seed = -50, epoch = 0) to the end (epoch = 50).

In (6), $\hat{e}_{sv}$ and $\hat{n}_{sv}$ represent the length of the "East" and "North" components of the satellite position, respectively, in the East-North-Up (ENU) coordinates relative to the receiver position.

E. Elevation angle

The quality of GNSS measurement is closely related to the elevation angle. If the satellite has a higher elevation angle, it is more likely to be blocked by buildings, resulting in NLOS signals. Therefore, we choose elevation angle as a feature for the learning of the ET neural network. As (7) shows, the elevation angle can be calculated from the GNSS raw data 7.

$$El = \arcsin\left(\frac{\hat{u}_{sv}}{\hat{l}}\right) \tag{7}$$

In (7), $\hat{u}_{sv}$ represents the upward component length of the satellite position in the ENS coordinates for the position of the GNSS receiver. l represents the distance between the satellite and the GNSS receiver. Although the exact location of the satellite is unknown, it is still possible to estimate the location of the satellite based on the ephemeris by combining the latitude and longitude of the GNSS receiver; finally, an elevation angle can be obtained within the error range.

IV. Experimental Results

This section analyzes the satellite visibility prediction results of the ET neural network from multiple angles through different binary classification evaluation indicators.

A. Experimental setup

The performance of the proposed ET neural network in predicting satellite visibility is verified using real data collected from the four places mentioned in the figure. The training portion of the network was trained for 50 epochs, and the test portion of the network took one-twentieth of all data for testing. The deep learning network was built and trained based on the Pytorch and Jupyter lab platforms.

B. ET network results

The satellite visibility prediction result is compared with the true label, as shown in Fig. 8. The left and right images in Fig. 8 show the comparison between the neural network prediction results and the label after 50 epochs of the training and untrained parts of the network, respectively (for

visualization, nine batches are randomly selected for observation). From the end of the parameter learning, this network has a good learning ability for two kinds of signals, rather than being biased toward only one kind of signal. At the same time, most of the learned probabilities tend to be 1 or 0, instead of 0.5, which indicates an excellent learning effect. Moreover, even if the signal changes rapidly between the two signals, the network can effectively distinguish between the two signals.

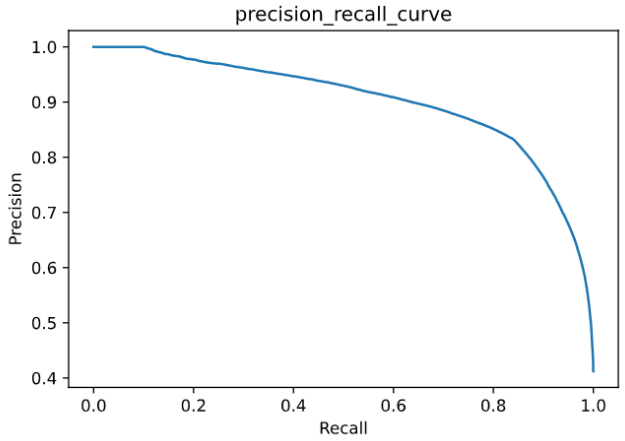


Fig. 9. P-R curve of satellite visibility prediction results.

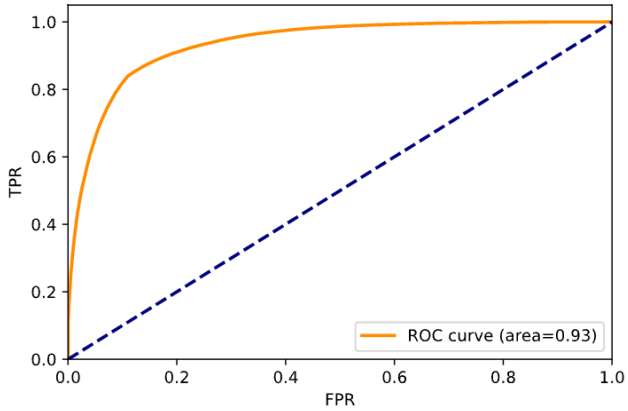


Fig. 10. ROC curve of satellite visibility prediction results.

The final satellite visibility prediction is determined in a binary form based on a threshold. In this study, we set the threshold to 0.5. This approach means that if the probability of satellite visibility prediction is over 0.5, it will be classified as NLOS; otherwise, it will be classified as LOS. The results of the satellite visibility prediction confusion matrix are shown in Table. The confusion matrix consists of four parts: true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The values of FN and FP are relatively low and relatively close, proving that the ET neural network is more accurate in judging positive and negative examples. After comparison with the true visibility label, we find that the results can reach an accuracy rate of 87.45% (an average of ten times that of the prediction results). The values of FN and FP are relatively low and relatively close, proving that the ET neural network is more accurate in judging positive and negative examples.

Precision refers to the proportion of samples that the model predicts to be positive. The recall rate refers to the proportion of samples that are predicted to be positive among the samples whose model label is positive. Expecting the two values to both be relatively high, we use the F1-score to represent the harmonic average of the two as the final evaluation index of the network's two-classification ability. It can be seen from the table that the precision rate is 84.43%, that the recall rate is 83.92%, and that the F1-Score is 84.18%. Therefore, it can be concluded that the ET neural network has relatively good classification ability in the F1-score view.

However, since the number of NLOS signals in the collected data is significantly less than the number of LOS signals, the accuracy rate is not sufficient to represent the classification ability of the ET neural network. Thus, we introduce the receiver operating characteristic (ROC) curve and the precision-recall (P-R) curve to represent the classification ability of the binary classification network, as shown in in Fig. 9–10. The horizontal axis of the P-R curve is the recall rate, and the vertical axis is the precision rate. For a point on the P-R curve, if the result is under that threshold the model judges it to be negative, and if the result is greater than the threshold, it is a positive sample. The horizontal axis of the ROC curve is the FPR, and the vertical axis is the TPR. The determination of ROC curve points is the same as that of the P-R curve. The AUC represents the area under the ROC curve, which characterizes the classifier's ability to rank positive samples in front of negative samples. It can be seen from Figs. 9-10 that even though the number of samples of NLOS and LOS differs, a smoother curve can still be obtained, proving that the ET neural network has a good classification ability under different thresholds.

TABLE I. CONFUSION MATRIX OF ET NEURAL NETWORK SATELLITE VISIBILITY PREDICTION.

Accuracy: 87.45%		Truth	
		NLOS (P)	LOS (N)
Prediction	NLOS(T)	33.37%(68917)	6.39% (13200)
	LOS(F)	6.15% (12705)	54.07%(111642)

C. *Contrast with other networks in feature extraction*

In [23], the superiority of ET neural networks over other networks was demonstrated. This experiment compares the ability of ET neural networks with traditional CNN and FCNNs+LSTM networks in extracting feature vectors. In [24], the CNN neural network used in the experiment was a five-layer one-dimensional convolutional network, and the network settings of the FCNNs+LSTM can be found in the literature [12]. We select the precision rate, recall rate, F1-score, and ROC curve as the evaluation indicators.

It can be seen from Fig. 11 and Table II that the various evaluation indicators of the ET neural network are higher than the other two neural networks. Next, to explain the experimental results, we theoretically analyze the superiority of the ET neural network. To visualize the comparison, we use a layer of CNN and self-attention to represent CNN and ET neural networks. First, as shown in Fig. 12, the CNN neural network uses kernels to characterize local features. However, for all the data, only one matrix with all the data information is obtained through the two-layer CNN.

TABLE II. COMPARISON OF PRECISION RATE, RECALL RATE, F1-SCORE RATE AND ACCURACY OF ET, CNN, AND FCNNS+LSTM NEURAL NETWORKS.

		Recall	Precision	F1-Score	Accuracy
CNN	LOS	88.54%	83.61%	86.00%	83.15%
	NLOS	75.56%	82.41%	78.83%	
FCNNs+LSTM	LOS	88.39%	87.54%	87.97%	85.43%
	NLOS	80.93%	82.15%	81.54%	
ET	LOS	90.47%	89.79%	**90.13%**	**87.45%**
	NLOS	83.92%	84.43%	84.18%	

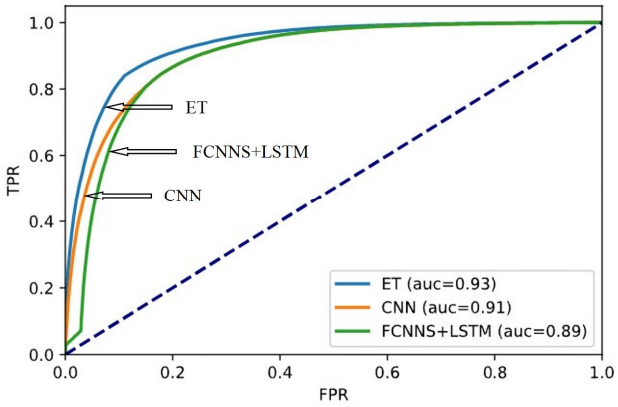


Fig. 11. Comparison of ROC curves of ET, CNN, and FCNNs+LSTM neural network.

However, compared with Fig. 13, five matrices with all the information can be obtained through a layer of self-attention. Moreover, because of the existence of the attention mechanism, the F1 corresponding to the feature vector of each dimension not only contains its information but also actively weights and receives the features of other dimensions.

Secondly, for the FCNN neural network, it is also possible to obtain the feature information of all data through one layer. However, the obtained matrix no longer has the information of the original corresponding dimension but has the same operation for each characteristic dimension. This is not conducive to further extraction of environmental information in the LSTM. Moreover, the multilayer FCNN leads to the neural network losing its nonlinear characteristics after too much linear fitting, which is not conducive to the final complex classification.

Therefore, both theoretically and experimentally, ET based on the attention mechanism can obtain better results in satellite visibility classification than other neural networks.

V. INVESTIGATION OF EACH PART OF THE ET NEURAL NETWORK

This section consists of four parts. The first two parts introduce the effects of different dimensions of feature vectors on classification and the relationship between them, as well as introduce what environmental features are extracted by environmental block 2 through attention visualization. The third part explains the adaptability of the ET neural network to different environments. The last part compares several different environment extraction methods to demonstrate the correctness of the ET neural networks.

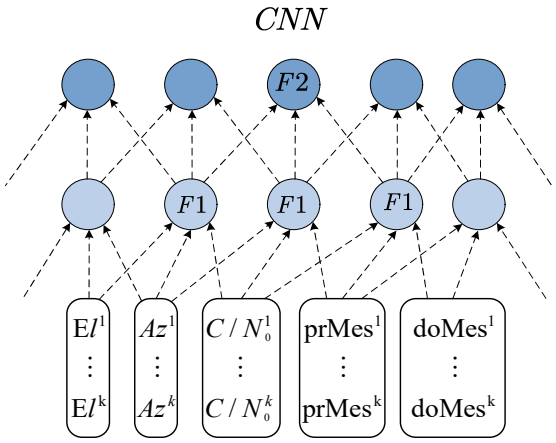


Fig. 12. CNN network feature extraction diagram.

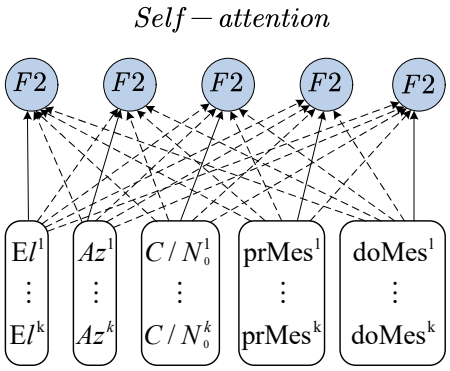


Fig. 13. Self-attention network feature extraction diagram.

A. Visualization of satellite visibility feature extraction

In this part, the effects of the different dimensions of feature vectors on GNSS signal visibility classification and the relationship between them are analyzed. As mentioned above, the encoder block trains twice in a loop. As shown in Fig.14, both the abscissa and the ordinate represent the different dimensions of the sample. For the four attention heads of the first circle, the attention weight is higher for three- to five-dimensional features. The corresponding dimensions are the carrier-to-noise density ratio, azimuth angle, and elevation angle. This result represents that the first

TABLE III. THE ABILITY OF THE ET NEURAL NETWORK TO PREDICT SATELLITE VISIBILITY UNDER DIFFERENT ENVIRONMENTAL CONDITIONS.

Datasets	Berlin Potsdamer Platz	Berlin Gendarme market	Frankfurt am Main Tower	Frankfurt am Main Westend Tower	All data (Without environmental block)	All data
Accuracy	86.03%	85.78%	87.36%	84.15%	84.35%	**87.45%**
F1-Score	83.10%	82.78%	85.37%	81.98%	81.36%	**84.18%**

circle of learning gives higher attention to the characteristics of these three dimensions. For the four attention heads in the second circle, it can be seen that their attention weights are higher in the second dimension. Thus, Doppler measurement is also important for feature extraction. However, its maximum attention weight does not exceed 0.5, meaning that the features of these five dimensions do not receive a great deal of network attention, but have a learning effect for each dimension. In summary, the network more or less extracts the response features from the feature vectors of these five dimensions and mainly extracts the carrier-to-noise density ratio, azimuth angle, and Doppler measurement.

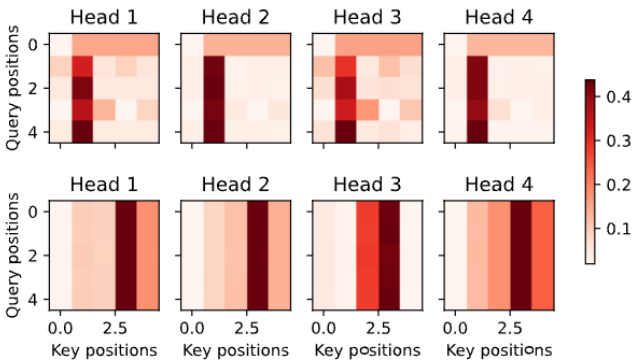


Fig.14. The attention weight of the multi-head attention is in the encoder block. Besides, the abscissa is a different key, and the ordinate is for different queries. The sum of the attention weights of all keys corresponding to each query is 1.

B. *Visualization of environmental feature extraction*

In this part, the necessity of setting up an environmental block from a visual point of view is explained. As shown in Fig. 15, for the multi-head attention mechanism adopted, each head learns different environmental characteristics. The abscissa represents the preliminary environmental features obtained after the environmental feature extraction of each feature vector. The ordinate represents the preliminary features after feature extraction for each feature vector. The shade of red in the figure represents the size of the attention weight. Generally, the image is in a diagonal form; that is, each preliminary feature pays more attention to its corresponding preliminary environment feature in the overall environment. The highest attention weight is no more than 15%, and it can be understood that each preliminary feature is not only limited to its corresponding preliminary environmental characteristics but also to a certain amount of learning for the surrounding environmental characteristics. Overall, this paper shows from the attention visualization that the network learns the surrounding environment of the signal collection site, and that this learning is both targeted and extensive.

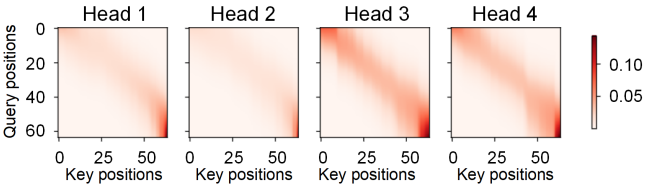


Fig.15. The attention weight of the multi-head attention is in environmental block 2. Besides, the abscissa is a different key, and the ordinate is for different queries. Moreover, the sum of the attention weights of all keys corresponding to each query is 1.

C. *The adaptability of the network to different environments*

Through experiments, it is found that simply extracting features from feature vectors causes some problems. As shown in Table III, columns 2 to 5 represent the results obtained through the ET neural network using the data obtained from the four different collection locations in Fig. 7. The sixth and seventh columns use the data of all four locations for training, and they either use or do not use the environmental block in Fig. 2, respectively. After conducting separate training on the data of each collection site, we found that the accuracy of the final model obtained from different collection sites was different. In addition, the classification accuracy obtained by separate training is greater than the classification accuracy obtained by data-combined training. Therefore, to increase the model's adaptability to the environment, we use the attention mechanism to extract environmental features. Ultimately, this method achieves good results.

D. *Comparison of different environmental feature extraction methods*

To prove that the ET neural network uses two layers of attention selection to be correct, we contrast it with the model a good Seq2seq, which has a strong ability of temporal feature extraction to extract environmental features. The LSTM and self-attention modules are used as feature extraction methods[23–26].

As shown in Table IV, it can be seen that the longitudinal use of the ET neural network has relatively good performance in both indicators. To explain the experimental results, we theoretically analyze the correctness of selecting ET nerves. To facilitate the analysis, we simplify the LSTM, as shown in Fig.16. In Fig.16, feature 2(F2) is obtained through k times of LSTM feature extraction, the result of each extraction is input into the next LSTM block, and the environment feature is finally obtained. However, for the features in a batch, their importance is the same. When the batch size is too large, the previous feature information will inevitably be weakened. Although the result can be optimized by weighting, the result is still not ideal. In contrast to the ET model, its principle is

TABLE IV. COMPARISON OF SEQ2SEQ (LSTM), SEQ2SEQ (ATTENTION), AND ET NEURAL NETWORKS IN ENVIRONMENTAL FEATURE EXTRACTION.

	Seq2seq (LSTM)	Seq2seq (Attention)	ET
Accuracy	84.21%	85.73%	**87.45%**
F1-Score	80.10%	83.49%	**84.18%**

shown in Fig. 12. Each feature vector can be balanced and actively extract information. In summary, theoretically, the ET model also has better environmental feature extraction capabilities.

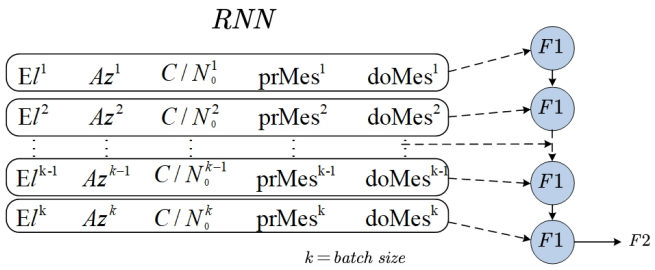


Fig.16. Recurrent neural network environment feature extraction diagram

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed an environmental transformer (ET) neural network based on the attention mechanism. The five-dimensional feature vector is extracted from the original data as the input of the network. The neural network as a whole consists of two parts. The main body is a satellite signal visibility classification network based on the encoder-decoder architecture. It can extract feature information hidden in feature vectors based on additive attention. Through the visual analysis of attention, we obtained the role and relationship of the input feature vector in the classification. Through experiments, we found that the performance of traditional classification models for data in different environments decreased. Therefore, to increase the model's ability to adapt to the environment, we introduced the environmental block. This allows the feature vector containing satellite visibility information to actively extract environmental information from the environmental block. To understand the situation of environmental feature extraction, we use attention visualization for analysis. Finally, by combining the two capabilities, ET can achieve an accuracy of 87.45% in the classification of satellite signal visibility.

In the process of data preprocessing, to reduce the influence of noise on ET network training, we determined that five-dimensional features can represent satellite signals as input. However, the use of five-dimensional features is far from sufficient to represent the visibility information of the satellite, and some features that are helpful for classification are also lost in the process of eliminating noise. Therefore, to maximize the use of the satellite's visibility feature information, we are more likely to achieve better results using the original signal data as input. However, the current classification model that uses the original signal as input is affected by noise, and the classification result is always unsatisfactory. Therefore, the focus of our future research will be to reduce the impact of noise by optimizing the existing model or adapting the original data to better serve as the input feature.

Recently, multimodal learning has been rapidly developed. Therefore, we used the satellite signals received by the GNSS receiver as the network's classification input for signal visibility. For example, we can also use the picture information or video information of the surrounding environment as the input of the neural network. By training on a multimodal input, it may be possible to extract more signal visibility features in the environment, thereby obtaining better classification accuracy. Therefore, using multimodal learning to simultaneously train different sources of information is also an important direction for future research.

ACKNOWLEDGMENT

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

REFERENCES

[1] Y.-T. Chan, W.-Y. Tsui, H.-C. So, and P. Ching, "Time-of-arrival based localization under NLOS conditions," IEEE Trans. Veh. Technol., vol. 55, no. 1, pp. 17–24, Jan. 2006.
[2] I. Guvenc and C.-C. Chong, "A Survey on TOA Based Wireless Localization and NLOS Mitigation Techniques," IEEE Commun. Surv. Tutor., vol. 11, no. 3, pp. 107–124, 2009.
[3] L.-T. Hsu, "Analysis and modeling GPS NLOS effect in a highly urbanized area," GPS Solut., vol. 22, no. 1, p. 7, Jan. 2018.
[4] R. Sun, G. Wang, W. Zhang, L.-T. Hsu, and W. Y. Ochieng, "A gradient boosting decision tree based GPS signal reception classification algorithm," Appl. Soft Comput.
[5] R. Sun et al., "Improving GPS Code Phase Positioning Accuracy in Urban Environments Using Machine Learning," IEEE Internet Things J., vol. 8, no. 8, pp. 7065–7078, Apr. 2021.
[6] L.-T. Hsu, "GNSS multipath detection using a machine learning approach," in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Oct. 2017, pp. 1–6.
[7] B. Xu, Q. Jia, Y. Luo, and L.-T. Hsu, "Intelligent GPS L1 LOS/Multipath/NLOS Classifiers Based on Correlator-, RINEX- and NMEA-Level Measurements," Remote Sens., vol. 11, no. 16, Art. no. 16, Jan. 2019.
[8] T. Suzuki and Y. Amano, "NLOS Multipath Classification of GNSS Signal Correlation Output Using Machine Learning," Sensors, vol. 21, no. 7, p. 2503, Apr. 2021.
[9] H. Xu, A. Angrisano, S. Gaglione, and L.-T. Hsu, "Machine learning based LOS/NLOS classifier and robust estimator for GNSS shadow matching," Satell. Navig., vol. 1, no. 1, p. 15, Dec. 2020.
[10] J.-S. Choi, W.-H. Lee, J.-H. Lee, J.-H. Lee, and S.-C. Kim, "Deep Learning Based NLOS Identification With Commodity WLAN Devices," IEEE Trans. Veh. Technol. vol. 67, no. 4, pp. 3295–3303, Apr. 2018.
[11] C. Jiang, J. Shen, S. Chen, Y. Chen, D. Liu, and Y. Bo, "UWB NLOS/LOS Classification Using Deep Learning Method," IEEE Commun. Lett., vol. 24, no. 10, pp. 2226–2230, Oct. 2020.
[12] G. Zhang, P. Xu, H. Xu, and L.-T. Hsu, "Prediction on the Urban GNSS Measurement Uncertainty Based on Deep Learning Networks With Long Short-Term Memory," IEEE Sens. J., vol. 21, no. 18, pp. 20563–20577, Sep. 2021.
[13] J. Park, S. Nam, H. Choi, Y. Ko, and Y.-B. Ko, "Improving Deep Learning-Based UWB LOS/NLOS Identification with Transfer Learning: An Empirical Approach," Electronics, vol. 9, no. 10, Art. no. 10, Oct. 2020.

[14] A. Vaswani et al., “Attention is All you Need,” in Advances in Neural Information Processing Systems, 2017, vol. 30. Accessed: Nov .04, 2021.
[15] K. Cho et al., “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” ArXiv14061078 Cs Stat, Sep. 2014, Accessed: Nov .05, 2021.
[16] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” ArXiv Prepr. ArXiv14090473, 2014.
[17] Z. Lin et al., “A structured self-attentive sentence embedding,” ArXiv Prepr. ArXiv170303130, 2017.
[18] A. P. Parikh, O. Täckström, D. Das, and J. Uszkoreit, “A decomposable attention model for natural language inference,” ArXiv Prepr. ArXiv160601933, 2016.
[19] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” J. Mach. Learn. Res., vol. 15, no. 1, pp. 1929–1958, 2014.
[21] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer Normalization,” ArXiv160706450 Cs Stat, Jul. 2016, Accessed: Nov .05, 2021.
[22] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” ArXiv Prepr. ArXiv14126980, 2014.
[23] J. Cheng, L. Dong, and M. Lapata, “Long Short-Term Memory-Networks for Machine Reading,” ArXiv160106733 Cs, Sep. 2016, Accessed: Nov .05, 2021.
[24] T. Fawcett, “An introduction to ROC analysis,” Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, Jun. 2006.
[25] M.-T. Luong, H. Pham, and C. D. Manning, “Effective Approaches to Attention-based Neural Machine Translation,” ArXiv150804025 Cs, Sep. 2015, Accessed: Nov .06, 2021.
[26] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in Advances in neural information processing systems, 2014, pp. 3104–3112.