Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/88/20241739

Research on fiber OTDR signal recognition and classification based on the ViT model

Qinhua Zhang

School of Communication & Information Engineering, Shanghai University, Shanghai, 200444, China

zhangqinzhe2012@126.com

Abstract. In the field of event recognition for phase-sensitive optical time-domain reflectometry (Φ-OTDR), convolutional neural networks (CNNs) have been the mainstream tool. However, Transformer models, with their self-attention mechanism, have provided a new perspective for image recognition tasks. This paper proposes a Φ-OTDR event recognition method based on the Transformer model and experimentally demonstrates its significant advantages over CNNs. The method utilizes an attention-based feature extraction approach to recognize distributed fiber optic sensing signals in a phase-sensitive optical time-domain reflectometer (φ-OTDR). In perimeter defense applications, various interferences and noise severely affect the recognition of φ-OTDR sensing signals, leading to false alarms, making accurate signal recognition both challenging and necessary. Extracting signal features and using machine learning classification models to recognize signals has been a research hotspot in recent years. The effectiveness of feature extraction is crucial to classification performance. This project introduces a temporal-spatial dual attention mechanism into the feature extraction of fiber optic sensing signals to improve the accuracy and robustness of signal recognition.

Keywords: OTDR fiber optic signal, Vision Transformer (ViT), signal recognition, machine learning.

1. Introduction

The widely regarded φ-OTDR system has several advantages, such as a wide frequency range and high spatial resolution. However, in perimeter defense applications, φ-OTDR signals typically contain a lot of noise and spurious signals, which can easily cause false alarms. Therefore, it is necessary to identify the signal types for accurate early warning.

Extracting signal features and using machine learning classification models for signal recognition has been a research hotspot in recent years. In signal recognition, feature extraction is crucial. By selecting appropriate methods to extract key features from the raw signals across different domains, such as statistical, frequency domain, time domain, and scale transform features, these methods can complement each other to reveal the distribution, frequency, and temporal information of the signals, thereby improving the ability to recognize events. The attention mechanism can automatically focus on key features during the process of extracting features from sensing signals, enhancing the accuracy and robustness of signal recognition.

DOI: 10.54254/2755-2721/88/20241739

Currently, more and more studies are incorporating attention mechanisms into signal recognition models. For example, a temporal attention mechanism has been introduced in convolutional neural networks (CNNs), channel and spatial attention mechanisms have been introduced in knowledge distillation models, temporal and spatial attention mechanisms have been added to temporal convolutional network (TCN) models and bidirectional long short-term memory neural networks (BiLSTM), and a statistical feature attention mechanism has been incorporated into LSTM models, among others [1-4]. These methods have further optimized the classification performance of the models, with recognition rates improving to varying degrees, with a maximum improvement of 4.1%. However, there is still room for improvement in recognition accuracy and the number of data types. This study focuses on the in-depth analysis and signal recognition of spectrograms obtained using the Vision Transformer (ViT) model based on coherent optical time-domain reflectometry technology. As an advanced fiber optic detection method, phase-sensitive OTDR technology provides rich information about the fiber state, but the accurate recognition and classification of its signals play a crucial role in perimeter defense applications.

By introducing the ViT model, this study aims to overcome the limitations of traditional convolutional neural networks in processing one-dimensional time-series data. It utilizes the self-attention mechanism of the ViT model to capture long-range dependencies in spectrograms, thereby improving the accuracy and reliability of signal recognition.

The research results not only have academic value, providing new perspectives and methods for the application of deep learning in the field of signal processing, but also offer the industry an efficient tool for fiber optic signal recognition, helping to improve the accuracy and robustness of fiber optic signal identification.

2. Data Pre-processing

2.1. Dataset Overview

The dataset used in this project was provided by Zhejiang University and was collected using phase-sensitive OTDR fiber sensors [5]. After processing, the data covers eight different categories, including noise, climbing, descending, knocking, shaking, rubbing, walking, and PCSV. This dataset provides researchers with a rich resource. Each category's images represent the signal variations of the fiber optic signals under the influence of different behaviors.

2.2. Sharpening

The authors encountered some challenges when attempting to improve the model's accuracy in the fiber optic signal spectrogram recognition project through image sharpening. Although sharpening can enhance the edges and details of an image and improve its clarity, this method did not bring the expected improvements in this project. Possible reasons include over-sharpening, which led to unnatural edges and artifacts in the images, interfering with the model's recognition of signal features. The specific textures and patterns in the dataset might require more complex image processing techniques [6]. The ViT model could be overly sensitive to minor changes in image features, and sharpening altered these features. Additionally, sharpening might have introduced extra noise, obscuring important signals in the images. Furthermore, there might be more suitable preprocessing methods, or the training strategy of the model itself might need adjustment. These factors combined resulted in the sharpening process not effectively improving the model's performance, with the final validation accuracy only reaching 85%.

2.3. Contrast Adjustment

By adjusting the contrast, this study aimed to enhance feature extraction, but the results were not ideal. This may be because the original contrast of the spectrograms was already sufficient, so further enhancing the contrast did not significantly help improve the model's performance. Additionally, if the adjustment is too extreme, it may introduce additional noise or distortion, which could interfere with the model's accurate recognition of signal features. In some cases, contrast adjustment might require more precise control or need to be combined with other image processing techniques. As a result, the final validation accuracy was only 86% [7].

2.4. Binarization
In an attempt to improve the model's accuracy in the fiber optic signal spectrogram recognition project, the authors chose binarization as a method. Binarization simplifies the image by converting it to black and white, with the goal of highlighting key features. However, this attempt was unsuccessful and led to a decrease in accuracy. The likely reason is that the binarization process overly simplified the image, resulting in the loss of important signal details and grayscale variations, which are crucial for the model to distinguish between different categories.

2.5. Denoising
In the project for image recognition of fiber optic signal spectrograms, it was observed from the confusion matrix that many images were incorrectly classified as noise categories. This suggests that there might be residual noise or disturbances in the dataset, which are affecting the model's classification accuracy. Based on this observation, the authors implemented further denoising measures to reduce the negative impact of noise on the model's performance.

Denoising is an image preprocessing technique aimed at removing or reducing unwanted noise from an image. By applying denoising algorithms, random fluctuations and outliers in the image can be effectively cleared, preserving key information in the image [8-9]. In this project, denoising significantly improved the model's ability to distinguish between different categories, reduced misclassifications, and increased classification accuracy.

Ultimately, the denoised dataset led to better performance of the model during training and validation, as shown in Figure 2, with a test accuracy reaching high levels. This validated the hypothesis that noise in the original dataset was one of the major causes of classification errors. By reducing the noise, the model was able to more accurately recognize and classify fiber optic signals, achieving satisfactory results. This experience also underscores the importance of data preprocessing in image recognition tasks and the necessity of ensuring data quality before model training.

3. Basic Structure of the Model

3.1. Model Architecture

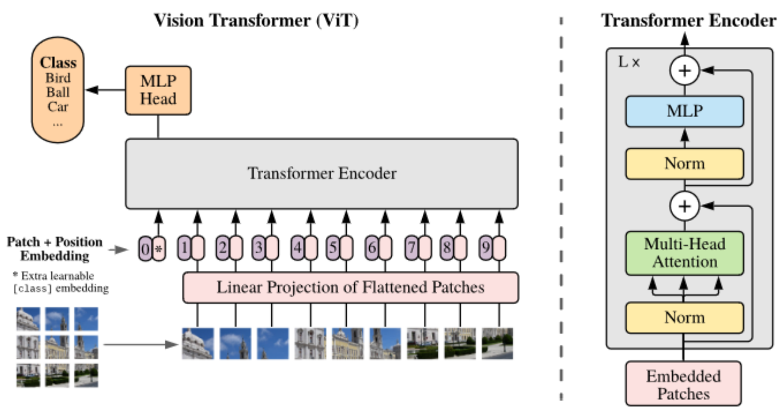

Figure 1. the diagram of ViT

As shown in Figure 1, the ViT model builds upon the Transformer architecture by adding an Embedding layer and an MLP classification layer to adapt to the image classification task of this project. For the

DOI: 10.54254/2755-2721/88/20241739

standard Transformer module, the input is a sequence of tokens (vectors), which is a two-dimensional matrix of [num_token, token_dim].

3.2. Embedding Layer

The data format for image data is [H, W, C], which is a three-dimensional matrix and not directly compatible with the Transformer's requirements. Therefore, an Embedding layer is used to transform the data. First, an image is divided into a set of patches of a given size. For example, in ViT-B/16, an input image of size (224x224) is divided into 16x16 patches, resulting in 196 patches. Each patch is then mapped to a one-dimensional vector through a linear transformation. For ViT-B/16, each patch with a shape of [16, 16, 3] is mapped to a vector of length 768 (referred to as a token). For instance, [16, 16, 3] -> [768][10].

3.3. Head MLP

The Head MLP is located at the end of the model and is responsible for converting the output from the feature extraction layers of the Transformer model into final classification results. In the ViT-B/16 model, the input and output shapes remain unchanged, meaning the model can handle data of the same dimensions. After feature extraction, the Head MLP's role is to extract global information representing the entire image or sequence, usually achieved through the [class] token, which captures the overall features of the input data. The Head MLP then maps this global feature vector to a lower-dimensional output space suitable for classification tasks. In this way, the model can generate a predicted class label for each input sample. In summary, the Head MLP acts as a bridge between high-dimensional features and classification decisions, ensuring the model can accurately recognize and classify various input data.

3.4. Results

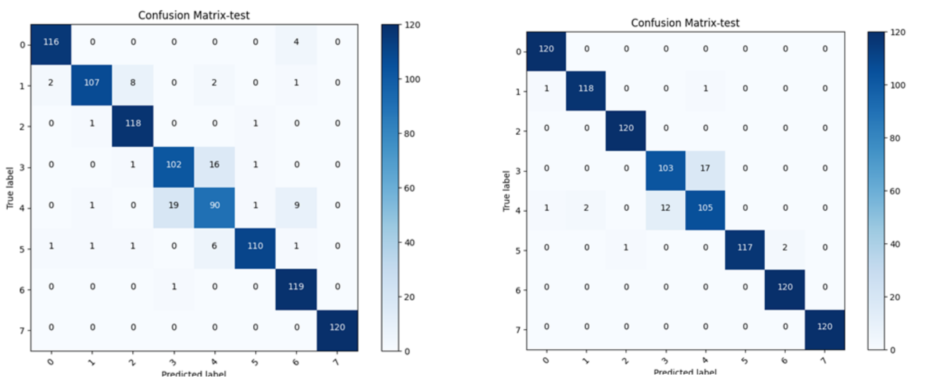

Figure 2. Confusion Matrix of the Original Dataset (Left) and Confusion Matrix of the Denoised Dataset (Right)

As shown in Figure 2, the model's classification accuracy has significantly improved, from 93% to 96%, indicating excellent performance. However, despite the overall good performance, some misclassifications remain. There are substantial errors with label 3 (shake) and label 4 (climb). The model misclassified two samples into other categories, suggesting that the model may have difficulties recognizing these specific categories.

In subsequent processing, we will examine whether the features used by the model are sufficient to distinguish between different categories, especially those prone to confusion. It may be necessary to introduce additional features or transform existing ones to enhance the model's discriminative ability. Additionally, we plan to optimize the model through methods such as cross-validation and error analysis.

By carefully analyzing the confusion matrix and implementing the above suggestions, we can further improve the model's classification accuracy, reduce misclassifications, and thereby enhance the model's reliability and effectiveness in practical applications.

4. Conclusion

This paper addresses the key issue of OTDR fiber optic signal recognition by proposing an automated classification method based on the Vision Transformer (ViT) model. By effectively modeling OTDR signals and optimizing the ViT network, a classification accuracy of 96% was achieved on the test set, significantly outperforming traditional machine learning methods. This research provides a new and effective solution for the intelligent application of OTDR technology and has significant practical implications for enhancing the reliability maintenance of fiber optic networks.

Although this study has achieved promising results, there are still some limitations that need to be addressed. The research is based on a limited OTDR signal dataset and lacks broader practical application research and evaluation. Future work could expand the dataset and conduct in-depth analysis for different application scenarios. Additionally, this study focuses solely on known categories for classification and recognition. If categories not present in the training set are introduced, the model's performance may deteriorate. Future research will aim to improve the model architecture further to handle not only known fiber optic signal categories but also to manage new signal types, better aligning with practical applications.

References

[1] Wen Tingkun. Research on Fiber Vibration Detection and Classification Based on Φ-OTDR [D]. Zhejiang University, 2022. DOI: 10.27461/d.cnki.gzjdx.2021.001283
[2] Tian Manling. Research on Disturbance Signal Classification Based on φ-OTDR Fiber Sensing System [D]. Beijing Jiaotong University, 2022. DOI: 10.26944/d.cnki.gbfju.2022.003206
[3] Yang Zhengguo. Research on Distributed Fiber Sensing Event Recognition Methods Based on Deep Learning [D]. Qilu University of Technology, 2023. DOI: 10.27278/d.cnki.gsdq c.2023.000316
[4] Chen, X. and Xu, C., 2020. Disturbance pattern recognition based on an ALSTM in a long-distance φ-OTDR sensing system. Microwave and Optical Technology Letters, 62(1), pp.168-175.
[5] Pan, Y., Wen, T. and Ye, W., 2022. Time attention analysis method for vibration pattern recognition of distributed optic fiber sensor. Optik, 251, p.168127.
[6] Lu Peng, Li Jinli. Normal Histogram Matching and Non-Sharpening Masking for Infrared Image Enhancement [J/OL]. Radio Communication Technology, 1-8 [2024-07-29]. http://kns.cnki.net/kcms/detail/13.1099.TN.20240628.1130.004.html.
[7] Xiang Ling, Gao Xin, Wang Yan. Laplacian-Based Noise Text Image Binarization Variational Model [J]. Journal of Chongqing Normal University (Natural Science Edition), 2023, 40(06): 86-94.
[8] Liu Yuli, Wang Kechao, Liu Lin. Machine Learning-Based Denoising Method for Multispe ctral Blurred Images [J]. Laser Magazine, 2022, 43(05): 156-160. DOI: 10.14016/j.cnki.j gzz.2022.05.156.
[9] Jia Xiangyu. Wavelet Transform-Based Denoising and Enhancement Algorithm for Nighttime Low-Luminance Images [J]. Information Technology and Informationization, 2019, (02): 107-109.
[10] Xue C, Deng Z, Yang W, et al.Vision transformer-based robust learning for cloth-changing person re-identification[J]. Applied Soft Computing, 2024, 163111891-111891.