Anomaly Detection and Localization in Optical Networks Using Vision Transformer and SOP Monitoring

K. Abdelli[1], M. Lonardi[1], J. Gripp[2], D. Correa[2], S. Olsson[2], F. Boitier[1], and P. Layec[1]
1: Nokia Bell Labs, 2: Nokia USA
E-mail: khouloud.abdelli@nokia.com

Abstract: We introduce an innovative vision transformer approach to identify and precisely locate high-risk events, including fiber cut precursors, in state-of-polarization derived spectrograms. Our method achieves impressive 97% diagnostic accuracy and precise temporal localization (6-ms-RMSE).

1. Introduction

Continuous state of polarization (SOP) monitoring is crucial in optical networks for early disturbance detection, such as proactive fiber damage detection [1]. SOP data provides valuable insights into the polarization characteristics of light signals traversing the network. By continuously monitoring SOP and by harnessing the capabilities of machine learning (ML) algorithms, subtle anomalies, or changes in SOP patterns (deviations from the norm) can be spotted. These disturbances often serve as early warning signs of impending fiber cuts or network outages. By proactively identifying and addressing these issues before they escalate into network, operators can safeguard the integrity of their optical infrastructure, minimize downtime, and ensure uninterrupted data flow transmission.

Previous works examined the use of a naive Bayes classifier for identifying mechanical stress-induced vibrations (SOP transients) in optical fibers, which serve as precursors to fiber breaks [1,2]. These transients were induced by robotic arm movements on a fiber and measured by a coherent receiver. [3] proposed a transfer learning approach for risky event classification with very limited SOP data. While the efficiency of ML algorithms in detecting SOP disturbances has received substantial attention, an unexplored realm pertains to precisely timestamping SOP transients. Moreover, the utilization of ML algorithms to handle overlapping events, frequently encountered in practical field scenarios, remains an unexplored area of research.

In this work, we propose a vision transformer-based (ViT) approach for anomaly detection and localization (ADL), called ViT-ADL in the following, for simultaneously identifying and accurately locating SOP vibrations in both the temporal and spectral domains, even in the presence of overlapping events. Our method harnesses spectrograms derived from a continuous stream of SOP measurements. We validated its efficiency using experimental SOP data collected from a bidirectional 2600 km link. To our knowledge, this study is the first to explore ML techniques for the classification and localization of overlapping events, advancing beyond previous research on shorter links.

2. Anomaly Diagnosis with Vision Transformers on Spectrogram-Derived SOP Time Series

An SOP monitoring functionality incorporated into a coherent receiver is used to collect SOP fluctuations in real-time. Specifically, the Stokes parameters ($\boldsymbol{S_1}$, $\boldsymbol{S_2}$, and $\boldsymbol{S_3}$) are recorded periodically. Each SOP recording is a multivariate time series of variable length $\boldsymbol{T}$, that incorporates the variation of Stokes parameters as a function of time. The SOP recordings are then sent for analysis to a remote-control system, such as a software-defined networking (SDN) controller. Our approach consists of three steps: i) splitting the continuous SOP time series; ii) transforming time series into a single image made up with 3 stacked (1 per Stokes parameter) time-frequency spectrograms, iii) an ML model leveraging vision transformer to output the event type, start time $\boldsymbol{x_1}$, end time $\boldsymbol{x_2}$, and frequency range ($\boldsymbol{y_1}$, $\boldsymbol{y_2}$) coordinates in the spectrogram. The first step splits the time series of each Stokes parameter $\boldsymbol{S_i}$ into non-overlapping sequences of fixed length $\boldsymbol{l}$ where $\boldsymbol{l < T}$. The second step transforms each generated time series sequence into an image of time-frequency spectrogram by using the short time Fourier transform (STFT). The resulting 2D spectrograms are of dimension of 28×28. The spectrograms are overlayed and stacked to generate a single image with three channels, then fed into the ViT-ADL model (last step). The ML model outputs the event type, accompanied by its corresponding coordinates ($\boldsymbol{x_1, y_1, x_2, y_2}$) within the spectrogram (represented as a bounding box). The start time of the event can be further transformed into the actual location along the optical fiber based on the fiber characteristics and if both ends share the same path and maintain a synchronization. Note that the accuracy of timestamping can significantly influence the precision of location determination. Following the identification and localization of the disturbance, mitigation procedures such as traffic rerouting, or dispatching technicians to fiber breakage sites can be implemented.

The architecture of ViT-ADL is shown in Fig. 1. First, an input image $\mathbf{X_i}$ derived from converting time series of Stokes parameters into spectrograms and stacked them as described above is partitioned into non-overlapping patches. Each patch is viewed by ViT-ADL as a distinct token. The sequence of generated patches undergoes a linear projection

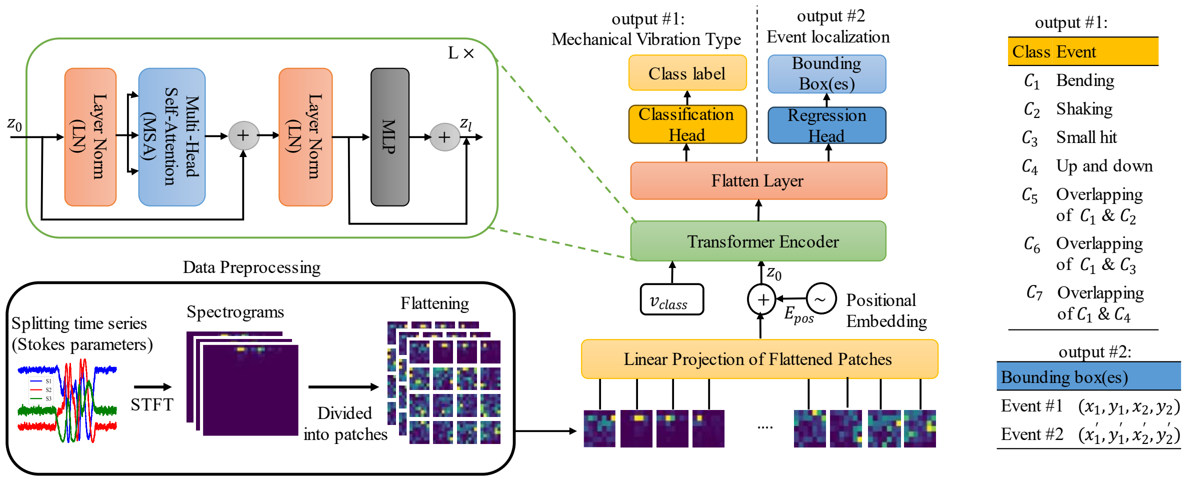

Fig.1: Proposed approach for disturbance detection and localization in spectrogram images derived from the Stokes parameter time series.

into a 1D vector with a model dimension of $\mathbf{d}$, achieved through a learned embedding matrix, before being fed into the Transformer encoder [4]. These embedded representations are subsequently concatenated together, incorporating a trainable classification token $\mathbf{v}_{\mathbf{class}}$ which is essential for carrying out the classification task. To preserve the spatial arrangement of the patches as observed in the original image, positional information is conserved through the inclusion of $\mathbf{E}_{\mathbf{pos}}$. The resulting sequence of embedded patches $\mathbf{z}_\mathbf{0}$ is passed to the Transformer encoder, which comprises $\mathbf{L}$ identical layers. Each layer contains a multiheaded self-attention (MSA) block, and a multilayer perceptron (MLP) block. Prior to each block, a normalization layer (LN) is introduced, and after each block, skip connections are applied. A Flatten layer is then applied to the final Transformer block's outputs to convert them into a one-dimensional representation. The flattened vector is simultaneously fed into both a classification head and a regression head. The classification head, which includes several fully connected layers followed by a Softmax activation function, outputs the type of event (class label). The regression head comprising of multiple fully connected layers with a linear activation function outputs the bounding box(es). The bounding box is defined by the event's start and end time coordinates, as well as the start and end frequency range coordinates. In this work, 7 event types namely $\mathbf{C_1}$: “bending”, $\mathbf{C_2}$: “shaking”, $\mathbf{C_3}$: “small hit”, $\mathbf{C_4}$: “up and down” and 3 overlapping events (i.e., $\mathbf{C_5}$: “bending and shaking”, $\mathbf{C_6}$: “bending and small hit”, $\mathbf{C_7}$: “bending and up and down”) are considered. For overlapping events, the regression head generates bounding boxes for each of the 2 events (i.e., $(\boldsymbol{x_1}, \boldsymbol{y_1}, \boldsymbol{x_2}, \boldsymbol{y_2})$, $(\boldsymbol{x'_1}, \boldsymbol{y'_1}, \boldsymbol{x'_2}, \boldsymbol{y'_2})$). The ViT-ADL’s loss is computed as a weighted sum of the regression head loss and the classification head loss.

3. Experimental Setup

The experimental setup shown in Fig. 2 is carried out to record SOP data for the validation of our approach. A link spanning 2600 kilometers, interconnecting different optical fibers through the utilization of ROADMs, is implemented by adopting a loop-back configuration through a ROADM port. To bolster signal integrity across this extensive distance, Optical Line Amplifiers (OLAs) are deployed to compensate signal attenuation. It is Raman-based amplifiers. An SOP monitoring functionality is seamlessly integrated into the coherent transponder to measure SOP. The SOP measurements are recorded in millisecond scale. To simulate mechanical vibrations along the optical fiber, including scenarios involving overlapping events, a pair of robots precisely controlled by Arduino microcontrollers is employed. The different movements or events induced by the robots encompass $C_{i\,\in[1..7]}$. To introduce an element of randomness, a polarization scrambler is harnessed to alter the initial state of polarization before each event generation. To generate a single event, one of the robots is employed individually. However, for the generation of overlapping events, the coordinated action of both robots is synchronized. The temporal gap between the onset of overlapping events is intentionally adjusted, ranging from 0.5 s to 4 s, to produce diverse patterns. In total, a dataset of 12,943 of samples (1849 for each investigated class) is built and used for the training and validation of our proposed approach.

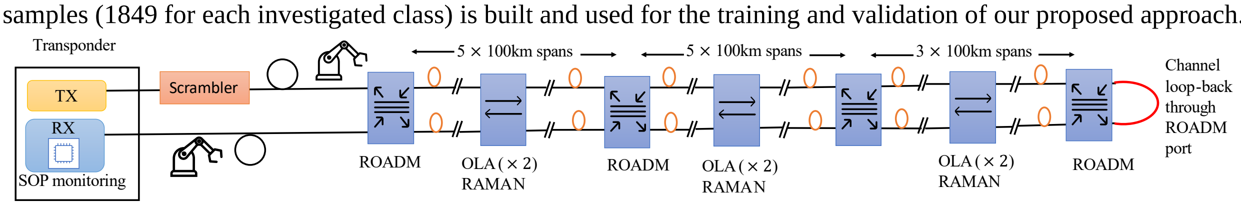

Fig. 2: Experimental setup for collecting SOP data including mechanical vibrations and overlapping events along the optical fiber.

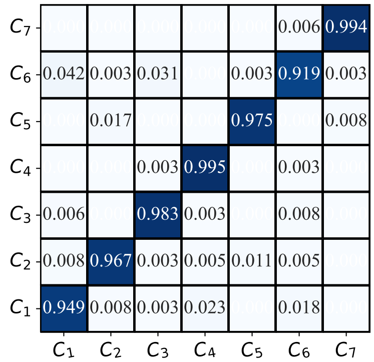

Fig.3: Confusion matrix with ViT-ADL.

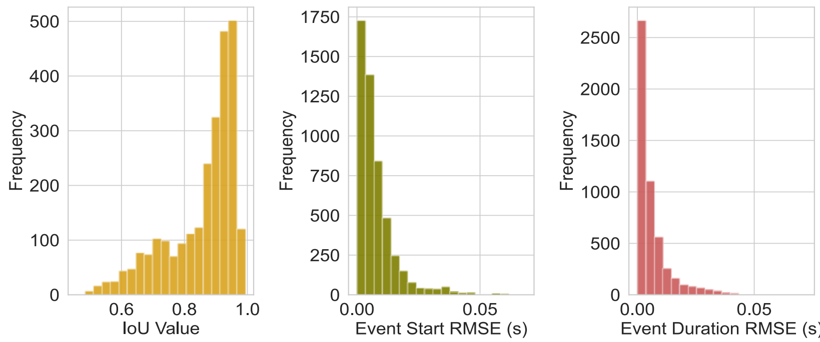

Fig. 4: Summary of localization performance yielded by our model: (a) histogram of IoU values, (b) histogram of RMSE of event start, and (c) histogram of RMSE of event duration.

4. Results and Discussion

The confusion matrix shown in Fig. 3, demonstrates that our model achieves an overall accuracy of 97%, each class exhibits an accuracy exceeding 91%. The results indicate that our approach excels at correctly identifying instances across all classes, even when events may overlap, with extremely few errors. The infrequent misclassification of the overlapping event C_6 can be attributed to the presence of C_1's patterns within C_6. The histogram of Intersection over Union (IoU), a metric quantifying the overlap between predicted and ground truth bounding boxes, as depicted in Fig. 4(a), demonstrates that our model achieves high IoU values, surpassing 0.5, with an average of 0.87, signifying its strong temporal localization performance. The histogram of root mean square error (RMSE) for event start time (Fig. 4 (b)) highlights our model's precision in event onset detection, consistently delivering minimal errors, with an average of just 6 ms. Additionally, the histogram of RMSE for event duration (Fig. 4 (c)) showcases our approach's accurate event duration prediction, with negligible errors, averaging a mere 7 ms. Fig. 5 presents illustrative outputs of our approach, demonstrating its efficiency. The bounding boxes predicted by our model closely align with the ground truth bounding boxes, as evidenced by the high IoU value, affirming the effectiveness of our approach. These results collectively underscore our model's precision in the temporal and spectral localization of events within spectrograms.

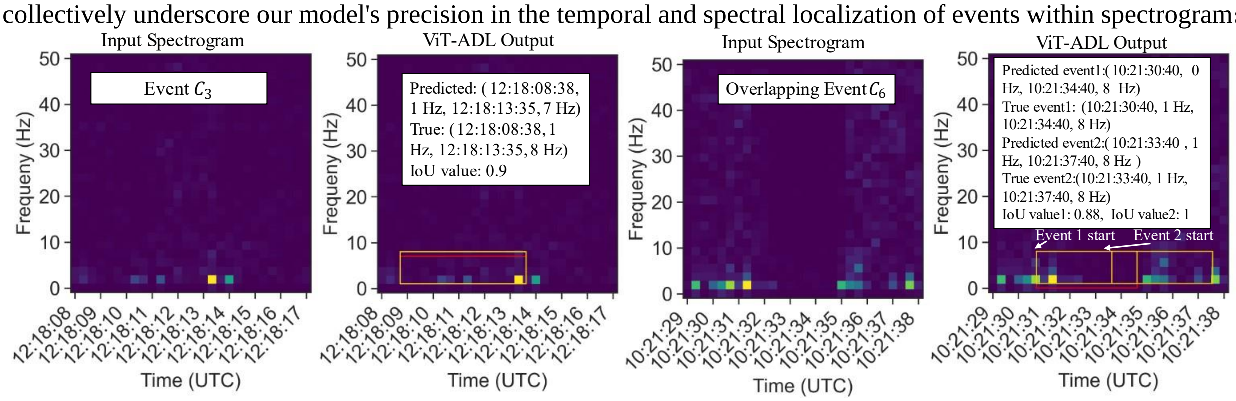

Fig. 5: Qualitive localization results. Notice the ground truth values are in gold. The predicted values are in **red**.

5. Conclusion

Our vision transformer approach achieved impressive results in mechanical vibration identification and simultaneous temporal and spectral localization within a monitored optical network. With a high diagnostic accuracy of 97% and an average event onset detection error of just 6 ms, along with an RMSE of event duration estimation at 7 ms, it holds the potential to significantly elevate network reliability and empower proactive maintenance planning.

This work has been supported by the French government through the Celtic-Next AINET-ANTILLAS research project.

6. References

[1] F. Boitier et al., "Proactive fiber damage detection in real-time coherent receiver", Proc. ECOC, pp. Th.2.F, 2017.
[2] V. Lemaire et al., "Proactive fiber break detection based on quaternion time series and automatic variable selection from relational data".
[3] K. Abdelli, et al., " Breaking Boundaries: Harnessing Unrelated Image Data for Robust Risky Event Classification with scarce State of Polarization Data," Proc. ECOC, 2023.
[4] V. Ashish et al. "Attention is All you Need." Proc. NIPS, 2017.