Anomaly Detection in Optical Spectra VIA Joint Optimization

ANOMALY DETECTION IN OPTICAL SPECTRA VIA JOINT OPTIMIZATION

Antonino Maria Rizzo[1], Luca Magri[1], Pietro Invernizzi[2], Enrico Sozio[2], Stefano Piciaccia[2], Alberto Tanzi[2], Stefano Binetti[2], Cesare Alippi[1,3], Giacomo Boracchi[1]

[1]Politecnico di Milano, [2]Cisco Photonics, [3]Università della Svizzera Italiana

ABSTRACT

Despite the remarkable progress of fiber optics in communication, little attention has been devoted to the automatic detection of anomalies in optical spectra, i.e., poorly transmitted channels. This task is typically addressed by ad-hoc heuristics that fall short in spectra presenting heavy distortions caused by optical amplifiers during transmission. We propose a method based on a joint optimization procedure for estimating the major trends that characterize the spectrum, enabling the detection of anomalies even in the presence of few channels and heavy distortions. Our experiments have shown that the proposed method can successfully localize anomalies achieving more than 98% accuracy, outperforming all competitors.

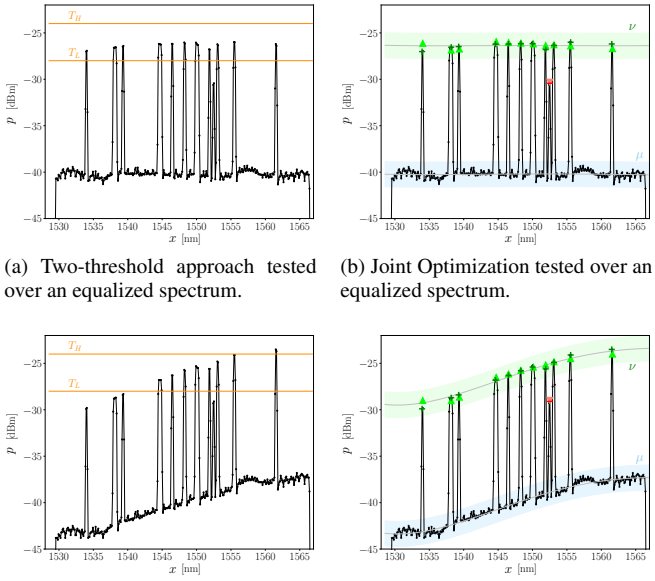
***Index Terms*—** Anomaly Detection, Optical Spectra, Joint Optimization.

1. INTRODUCTION

Fiber optics are essential to transmit data over long distances. User information is transmitted through channels that form an optical signal called *spectrum*. An example of such a signal is shown in Fig. 1a, where channels appear as well-equalized peaks, meaning that they have the same power level. At the bottom of the spectrum, the Amplified Spontaneous Emission (ASE), introduced by the amplifiers, provides a base power level in regions with no channels.

A significant problem that occurs in the composition of spectra is the presence of *anomalies*, which are mostly caused by faults in the fiber and typically result in non-equalized peaks (marked in Fig. 1b by the red triangle with the tip down). To guarantee proper data transmission, it is essential to design automatic procedures to detect anomalies to trigger corrective actions. Otherwise, systems mounted along fiber links that autonomously operate to improve the transmission quality (e.g., optical amplifiers) might undertake wrong actions leading to potentially dangerous cascade-effects.

Although anomaly detection is an active research field [1, 2], this problem has not been extensively investigated in the context of optical spectra. In practice, the common solution adopted in industrial applications consists of setting two thresholds on channel power, T_L and T_H (orange lines in Fig. 1a), to recognize as anomalies all the peaks that fall outside these two. However, this solution fails in the general case and is only effective on well-equalized spectra, which rarely happens in practice. As a matter of fact, when transmitted over long distances, amplifiers introduce fluctuations and tilts that distort the spectrum such that both channels and ASE follow a trend, denoted as ν and μ in Fig. 1d. When these trends are non-horizontal, the two-thresholds approach fails, as illustrated in Fig. 1c, where the spectrum is tilted, and the trend is no longer straight as in Fig. 1a. Since this anomaly has a higher power level than the leftmost channel, it is not possible to adjust the two thresholds to detect the anomaly without invalidating at least one channel.



(a) Two-threshold approach tested over an equalized spectrum.

(b) Joint Optimization tested over an equalized spectrum.

(c) Two-threshold approach tested over a distorted spectrum.

(d) Joint Optimization tested over a distorted spectrum.

Fig. 1: In (a) all the channels appear as peaks in the spectrum having the same power level (equalized), whereas in (c) the spectrum becomes distorted and the channels are no more equalized. While in (a) the thresholds T_H and T_L (colored orange) can successfully separate channels from anomalies, in (c) the use of thresholds becomes ineffective. Figures (b) and (d) depict the channel and ASE trends, ν and μ, estimated by our proposed method.

Anomalies could be in principle detected by considering the spectrum as a time series and using a supervised detection network to identify channels and anomalies. This strategy adapts visual recognition models, such as the Faster R-CNN [3] and YOLO [4], to 1D signals. Among the most recent works, this has been successfully applied to the detection of anomalies in ECG tracings [5], earthquakes in seismic waves [6], optical events in OTDR traces [7], and speech objects in a fixed-length audio signal [8]. Unfortunately, while localizing channels and anomalies is straightforward, classifying between them is difficult since anomalies are rare and have the same shape as channels, the only difference is that their power levels deviate from the channel trend.

In this work, we consider the general scenario where spectra are tilted and might follow a trend as in Fig. 1. We define anomalies as peaks that do not follow the channel trend. Therefore, instead of directly localizing channels and anomalies, our goal is to estimate the channel trend and detect anomalies as peaks departing from it.

To fit the channel trend, we resort to robust fitting techniques, such as RanSaC [9], an iterative algorithm that fits analytical models over randomly picked samples. Then, RanSaC assesses the models based on their consensus, namely the number of samples that lie within a user-specified tolerance ε, termed *inlier threshold*. The model having the highest consensus is then selected. Unfortunately, when few channels exist, the consensus can be influenced by anomalies, resulting in an utterly incorrect estimate. An example is shown in Fig. 4f, where, due to the scarcity of channels, the anomaly has been selected in the fitting, resulting in the wrong trend.

Our method overcomes this limitation in an unsupervised manner by leveraging a joint optimization procedure and exploiting RanSaC to robustly initialize the channel and ASE trends. In particular, since the fiber distortions alter the channel and ASE trends in a similar way, as shown in Fig. 1b and Fig. 1d, the joint optimization accounts for their similarity, enabling a better estimate of the channel trend and more accurate anomaly detection. In addition, the estimate of the channel trend is particularly beneficial since it can be used to detrend the spectrum, thus reducing transmission distortions.

Qualitative experiments performed on real-world spectra gathered from CISCO have shown that our method is accurate and capable of detecting anomalies even in the presence of severe distortions and tilts. A sample result is shown in Fig. 3. Quantitative analysis on a dataset of synthetically generated spectra, modeling the main distortions that occur in the fiber, confirms the advantages of our approach, as we achieve more than 98% accuracy.

2. PROBLEM FORMULATION

An optical spectrum $S = \{S(x_1), S(x_2), \ldots, S(x_N)\}$ can be modeled as a vector where x_i and $S(x_i)$ are the frequency and power of the *i*-th sample. Each spectrum S contains an unknown number of channels that can occur at any location. Each channel is characterized by a triplet (c_j, b_j, p_j) representing the channel's central frequency, bandwidth, and power, respectively.

At the bottom of the spectrum, there is a visible base power level termed the ASE. Channels and ASE follow two unknown trends, ν and μ, which we assume can be well approximated by polynomials, depicted in Fig. 1b in green and blue, respectively. Channels that do not conform to the channel trend ν are termed anomalies. Anomaly detection consists in automatically localizing the anomalies in the spectrum, if any, and determining their central frequency c_k, without any form of supervision.

3. PROPOSED METHOD

Our method aims to detect anomalies by robustly fitting the channel trend and identifying peaks that deviate from the trend as anomalies. We achieve accurate estimation of the channel trend by leveraging its similarity to the ASE trend. Our method is described in alg. 1. At a high-level, it consists of three stages: *i)* the initialization where we cluster candidate channels and ASE samples in two separate groups and coarsely fit the channel and ASE trends, ν and μ; *ii)* the joint optimization where we refine ν and μ to accurately follow the trend of the channels and ASE samples while promoting their similarity; *iii)* anomaly detection, where channels that deviate from the previously estimated channel trend ν are recognized as anomalies. These stages of the algorithm are further described below.

Algorithm 1 Joint Optimization

Input: Spectrum S, inlier threshold ε
Output: Central frequencies c_k of the K anomalies

Stage 1: Initialization.
1: Fit a linear trend $\ell \leftarrow$ LO-RanSaC$(x_i, S(x_i), \varepsilon)$
2: $h_r \leftarrow$ Construct the histogram of residuals.
3: $\mathcal{C}, \mathcal{N} \leftarrow \mathrm{Otsu}(h_r)$.
4: $\mathcal{C} \leftarrow \mathrm{FindPeaks}(\mathcal{C})$.
5: $\nu \leftarrow$ Fit polynomial to the samples in $\mathcal{C}$.
6: $\mu \leftarrow$ Fit polynomial to the samples in $\mathcal{N}$.
Stage 2: Joint optimization;
7: $\nu^*, \mu^* \leftarrow$ Jointly estimate trends as in Eq. (1).
Stage 3: Anomaly detection;
8: $c_k \leftarrow$ Recognize as anomalies $\{c_k \in \mathcal{C} : \mathrm{err}(p_k, \nu) > \varepsilon\}$.

3.1. Channel and ASE trends initialization

In the initialization step, we fit a linear trend ℓ to coarsely cluster channel and ASE samples. In particular, we fit analytical models (alg. 1, line 1) by leveraging LO-RanSaC [10] with a loose inlier threshold of 3.0 dB, thus including the spectra fluctuations in the inlier band. We compute the optimal separation threshold between channel and ASE samples by exploiting the Otsu method (alg. 1, line 3) fed with the histogram of residuals of all the samples from the trend ℓ (alg. 1, line 2). As a result, we obtain two clusters of samples that we denote as $\mathcal{C}$ (putative channels) and $\mathcal{N}$ (noise). Then, we locate peaks in $\mathcal{C}$ (alg. 1, line 4) as these are the best candidate for channels and anomalies. Finally, we independently fit the channel trend $\nu(x) = \sum_{i=0}^{d} = m_i x^i$ to the samples in $\mathcal{C}$ (alg. 1, line 5) and the ASE trend $\mu(x) = \sum_{i=0}^{d} = n_i x^i$ to the samples in $\mathcal{N}$ (alg. 1, line 6), where m_i and n_i are the coefficients of the polynomials. Notice that the two polynomials have the same order d.

3.2. Joint optimization

Our method accounts for the similarity of the channel and ASE trends in a joint optimization procedure that provides a robust estimate of the channel trend ν, even in the presence of anomalies, and in spectra having few channels.

We refine the initial guesses of ν and μ, computed in Sec. 3.1, by solving a non-linear optimization problem using the Levenberg-Marquardt [11] algorithm (alg. 1, line 8). Specifically, we jointly optimize the two trends ν and μ by minimizing the following loss:

$$\arg\min_{\mu,\nu} \sum_{p\in\mathcal{C}} \mathrm{err}(p,\nu)^2 + \sum_{p\in\mathcal{N}} \mathrm{err}(p,\mu)^2 + \frac{\lambda}{d}\sum_{j=1}^{d}(n_j - m_j)^2 \quad (1)$$

where $\mathrm{err}(p, \nu)$ represents the distance (over the vertical axis) of the trend ν to samples $p \in \mathcal{C}$, and similarly $\mathrm{err}(p, \mu)$ represents the distance between samples $p \in \mathcal{N}$ and the trend μ. The first two terms of Eq. (1) represent the data fidelity of the trends to the samples, whereas the third term enforces the similarity of the two trends allowing ν and μ to be close to each other, except a translation along the vertical axis that is encoded in their zero-degree terms n_0 and m_0. The parameter $\frac{\lambda}{d}$ balances the data fidelity and the similarity terms. In particular, the division by d allows setting the same value of λ for different order d of the trends. Note that the values n_j e m_j are the trend coefficients optimized at each optimization iteration.

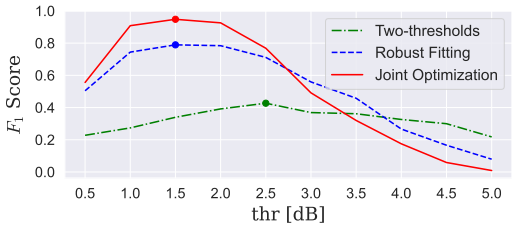


Fig. 2: F_1 Score vs. inlier thresholds. For each method, we selected the threshold value leading to the highest F1 Score. The two-threshold approach achieves a maximum F_1 Score of 0.42 at 2.5 dB. The robust fitting approach and our joint optimization use a threshold of 1.5 dB, achieving F_1 Scores of 0.79 and 0.95, respectively.

3.3. Anomaly Detection

Once the channel trend has been estimated by optimizing Eq. (1), detecting anomalies is straightforward. After having identified the channel trend ν, we recognize as anomalies the peaks $p \in \mathcal{C}$ that fall outside the inlier threshold ε of the channel trend, determining the green band illustrated in Fig. 1d.

4. EXPERIMENTS

This section is devoted to qualitative and quantitative testing of the method. Our experiments demonstrate that our joint optimization procedure enhances channel trend estimation, enabling the accurate detection of anomalies. Furthermore, while anomalies in spectra having few channels might fool robust fitting techniques, our method overcomes this problem.

4.1. Considered solutions

Two-thresholds. This solution consists in detecting anomalies by estimating all the peaks in S as they represent potential candidates for channels and anomalies. To this end, we employ the peak detection algorithm in [12]. The peaks are then classified as anomalies when they fall outside two user-defined thresholds, T_L and T_H, which is equivalent to selecting a single threshold T and an inlier threshold ε such that peaks falling outside the range $[T-\varepsilon, T+\varepsilon]$ are considered anomalies. We set the value T to the mean power of the identified channels for convenience. In all our experiments, the inlier threshold has been set to 2.5 dB, the value yielding the best detection performance in terms of F_1 Score, as in Fig. 2. Notice that this solution is the de-facto standard for industrial applications.

Robust Fitting. Robust techniques determine the trend followed by channels without being impaired by anomalies. To this purpose, we first detect peaks by [12] to detect candidate channels and anomalies. Then, the LO-RanSaC algorithm [10] fits a linear trend ℓ with the inlier threshold set to 1.5dB, which is the value allowing optimal performance as in Fig. 2. Anomalies are finally detected as the peaks that are outliers for the fitted linear trend. Notice that linear trends are less affected by noise and anomalies than higher-order polynomials that yield worse results.

Faster R-CNN. This is a supervised alternative that trains a 1D Faster R-CNN [3] on an annotated dataset of optical spectra to detect both channels and anomalies. The central frequency of the resulting detections is obtained by computing the center of the bounding boxes. The model has been trained with a learning rate of 10^{-3} with a decay of 0.5 each 15 epoch. The receptive field of the Feature Extractor is 278, and spectra were resampled so that the bandwidth of channels and anomalies is below 71 samples, allowing the network to cover a large enough portion of the spectrum in his prediction. The network has been trained on a dataset of 102 synthetic spectra exploiting horizontal shift and offset addition as the only forms of augmentation. A weighted loss has been used to mitigate class imbalance in the data.

Joint Optimization. This is our solution where we model the trends as polynomials of order $d = 4$. Higher-order polynomials would be too prone to overfitting and result in slower convergence of the algorithm. Our solution uses an inlier threshold of 1.5, the value for which it achieves the best performance, as in Fig. 2.

	Accuracy		Precision		Recall		F_1 Score	
	Mean	Var.	Mean	Var.	Mean	Var.	Mean	Var.
Two-thresholds	0.662	0.054	0.329	0.119	0.697	0.141	0.391	0.112
Robust Fitting	0.951	0.010	0.800	0.117	0.810	0.110	0.789	0.106
Faster R-CNN	0.857	0.006	0.000	0.000	0.000	0.000	0.000	0.000
Joint Optimization	**0.989**	0.002	**0.968**	0.028	**0.937**	0.036	**0.948**	0.030

Table 1: Quantitative results over the synthetic test set of optical spectra.

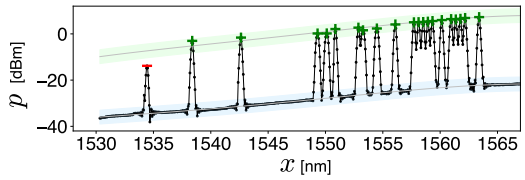


Fig. 3: Real-world spectrum analyzed with our joint optimization.

4.2. Dataset

We considered 95 real-world spectra acquired by CISCO in the field, an example is depicted Fig. 3. Large amounts of real-world spectra are difficult to gather as they might contain sensitive user information, and it is difficult to have annotated anomalies since most of them are acquired in a normal regime. Therefore, to have quantitative results, we have generated a synthetic dataset of 165 optical spectra. In order to establish a realistic benchmark, we carefully modeled the real fiber distortions, such as the fluctuations and tilt introduced by the amplifiers. Each spectrum contains from 1 to 50 channels, and of these, at most 5 anomalies. Overall, the dataset contains 2800 channels and 200 anomalies.

4.3. Figures of merit

To assess all the aforementioned methods, we employ standard performance measures, such as accuracy, precision, recall, and F_1 score. We first match each estimated anomaly to the corresponding ground-truth using the Hungarian algorithm [13] that estimates the function σ that maps each central frequency c_i to its nearest estimate $\hat{c}_{\sigma(i)}$. An anomaly is successfully localized when the distance between $\hat{c}_{\sigma(i)}$ and c_i falls within the channel bandwidth:

$$c_i - \frac{b_i}{2} \leq \hat{c}_i \leq c_i + \frac{b_i}{2}. \tag{2}$$

Unmatched peaks in the ground-truth are marked as false negatives, whereas unmatched predictions are marked as false positives.

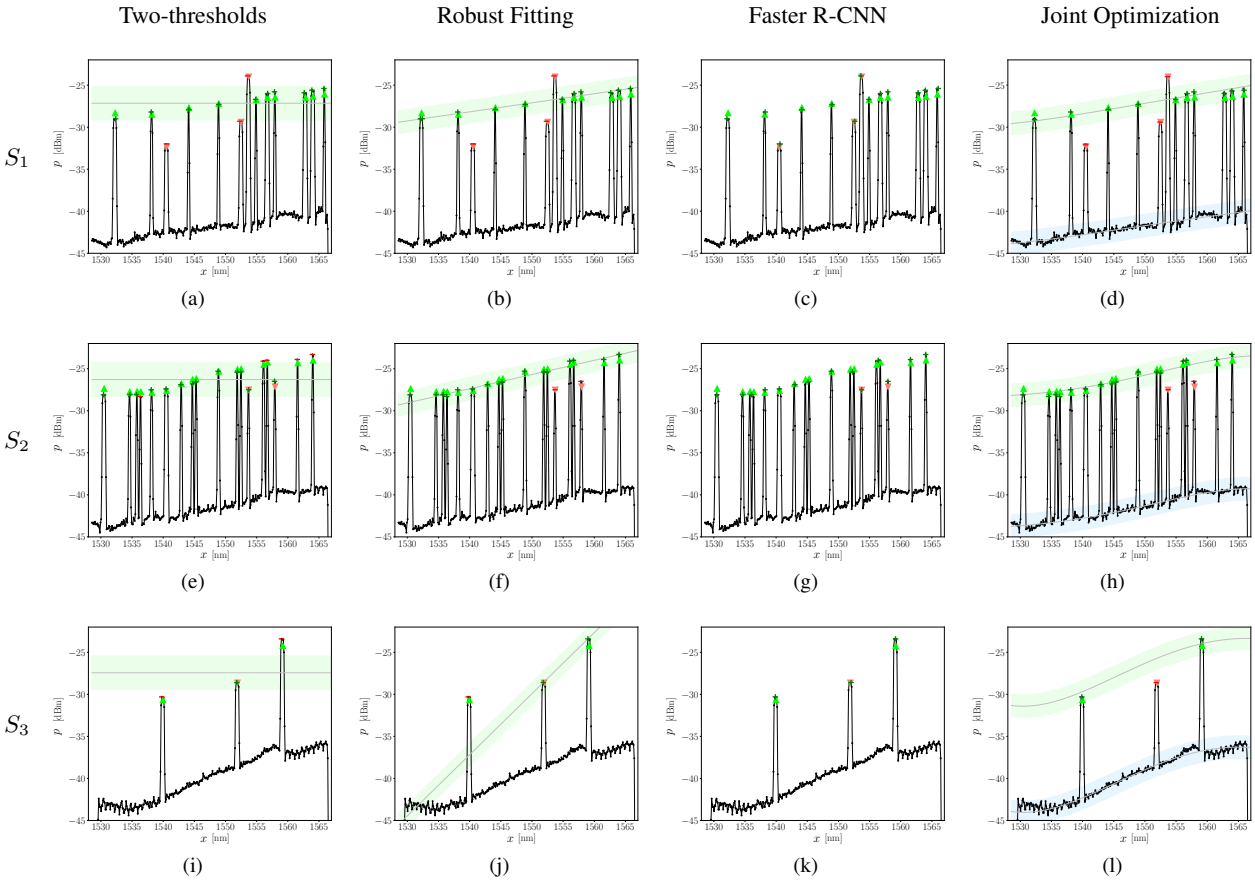


Fig. 4: The three synthetic spectra S_1, S_2 and S_3 have been analysed with the different methods. True channels are denoted with the green triangles with the tip up, whereas true anomalies are denoted with the red triangle with the tip down. Estimated channels are denoted with the green "$+$", whereas estimated anomalies are denoted with the red "$-$".

4.4. Discussion

The threshold of each method (except the Faster R-CNN which does not need one) has been determined by picking the values leading to the highest F_1 Score, as in Fig. 2. Results averaged over all the synthetic spectra are reported in Table 1 and indicate that our method is the best-performing algorithm according to all the selected metrics. Fig. 4 depicts three meaningful sample results.

The two-threshold approach achieves the best performance in spectrum S_1 having a nearly horizontal trend (Fig. 4a), but is ineffective on the tilted spectra S_2 and S_3 (Fig. 4e and Fig. 4i). Notice that the channels on the right-hand side of the spectrum are recognized as anomalies, as their peak power falls outside the green band.

The robust fitting approach handles both the flat and tilted spectra S_1 and S_2 (Fig. 4b and Fig. 4f), but it is ineffective in the presence of few channels, as the case of spectrum S_3 (Fig. 4j) where it selected the anomaly to fit the channel trend, resulting in a completely incorrect estimate.

The Faster R-CNN identifies all peaks in S_1, S_2, and S_3 as channels (Fig. 4c, Fig. 4g and Fig. 4k) due to the severe class imbalance in network training caused by the rarity of anomalies. However, it localizes all channels that act as true negatives in the accuracy computation. Indeed, accuracy reaches 85.7%, whereas precision, recall, and F_1 Score are zero.

Our joint optimization method can successfully detect anomalies in the flat spectrum S_1 (Fig. 4d), as well as in the presence of heavy tilt (Fig. 4h) and few channels (Fig. 4l), represented by spectra S_1 and S_2. Leveraging the ASE trend in the optimization, our method achieves 98.9% accuracy and 94.8% F_1 Score. When applied to a real-world spectrum, such as the one depicted in Fig. 3, which is highly tilted but presents almost no fluctuations, our method accurately detects the anomalies and estimates the channel trend.

5. CONCLUSIONS AND FUTURE WORK

In this work, we present a novel method to perform anomaly detection in optical spectra. The success of our approach is due to a joint optimization procedure that allows us to fit the channel trend and extracting useful information encoded in the ASE.

We plan to extend our method to consider multiple spectra simultaneously, thus incorporating the temporal dimension. By analyzing the transmission history, we can identify the causes of anomalies and trigger appropriate countermeasures. For instance, while persistent anomalies might hinder transmission and necessitate human intervention, transient anomalies might indicate channels introduced during spectrum acquisition and requiring no intervention.

Acknowledgements We gratefully acknowledge the support of *NVIDIA Corporation* with the GPUs granted by the Academic Hardware Grant Program to Giacomo Boracchi and Luca Magri.

6. REFERENCES

[1] Varun Chandola, Arindam Banerjee, and Vipin Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, jul 2009.

[2] Rémi Domingues, Maurizio Filippone, Pietro Michiardi, and Jihane Zouaoui, "A comparative evaluation of outlier detection algorithms: experiments and analyses," *Pattern Recognition, Volume 74, February 2018*, 2017.

[3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[4] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[5] Özal Yıldırım, Paweł Pławiak, Ru-San Tan, and U. Rajendra Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ecg signals," *Computers in Biology and Medicine*, vol. 102, pp. 411–420, 2018.

[6] Yue Wu, Youzuo Lin, Zheng Zhou, David Chas Bolton, Ji Liu, and Paul Johnson, "Deepdetect: A cascaded region-based densely connected network for seismic event detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 62–75, 2018.

[7] Antonino Maria Rizzo, Luca Magri, Davide Rutigliano, Pietro Invernizzi, Enrico Sozio, Cesare Alippi, Stefano Binetti, and Giacomo Boracchi, "Known and unknown event detection in OTDR traces by deep learning networks," *Neural Computing and Applications*, 2022.

[8] Yael Segal, Tzeviya Sylvia Fuchs, and Joseph Keshet, "Speechyolo: Detection and localization of speech objects," *Proc. Interspeech 2019*, pp. 4210–4214, 2019.

[9] Martin A. Fischler and Robert C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[10] Ondřej Chum, Jiří Matas, and Josef Kittler, "Locally optimized ransac," in *Pattern Recognition*, Bernd Michaelis and Gerald Krell, Eds., Berlin, Heidelberg, 2003, pp. 236–243, Springer Berlin Heidelberg.

[11] Jorge J. Moré, "The levenberg-marquardt algorithm: Implementation and theory," in *Numerical Analysis*, G. A. Watson, Ed., Berlin, Heidelberg, 1978, pp. 105–116, Springer Berlin Heidelberg.

[12] Scipy, "Find peaks," `https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html`, Accessed: 2022-10-10.

[13] James Munkres, "Algorithms for the assignment and transportation problems," *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957.