Monitoring Data Augmentation of Spectral Information using VAE and GAN for Soft-Failure Identification

Lars E. Kruse*, Sebastian Kühl, Annika Dochhan and Stephan Pachnicke
Chair of Communications, Kiel University, 24143 Kiel, Germany
**E-mail: lars.kruse@tf.uni-kiel.de*

Abstract: We propose data augmentation of monitoring information using VAE and GAN to reduce the amount of required soft-failure training data. Results show that only 5 samples per failure type are needed for F1-scores above 0.9.

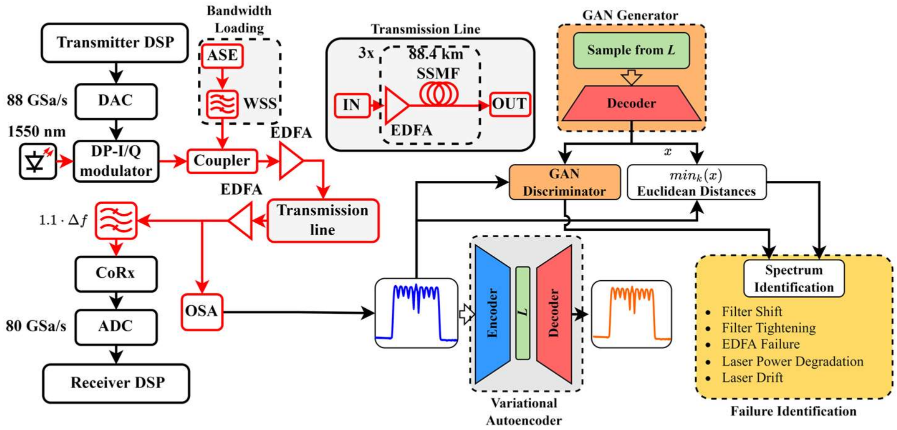
1. Introduction

The need for high speed data is growing rapidly in today's digital age. In our interconnected world, the loss of optical connectivity means both, data loss and service level agreements not being met. The implementation of machine learning algorithms is an increasingly promising approach for proactive maintenance in future networks, rather than relying on conservative design principles, guaranteed redundancies, and threshold-based fault detection alarms [1, 2]. However, most machine learning algorithms require a significant amount of training data to ensure reliable and accurate operation. To ensure the availability of such data, the role of optical performance monitoring (OPM) is crucial. In this regard, network-wide OPM is important for the training, validation, and development of machine learning algorithms for fault management. This may be achieved by the use of optical spectrum analyzers (OSAs) at strategic nodes throughout the network to capture the optical spectrum. The collected data will be employed afterwards in machine learning (ML) frameworks. Managing soft-failures, i.e., events that progressively degrade the transmission quality, is also extremely important as they deteriorate the transmission quality and may lead to disruptions of services, i.e., hard failures. Therefore, soft-failure management is increasingly shifting into the focus.

Recently, great efforts have been made by the community to find reliable ML algorithms to deal with soft-failures in optical networks. For example, in [3], the received power and the bit error ratio (BER) are used as input features for an ML-model to detect and identify soft-failures caused by signal overlapping, filter tightening and filter shift. In [4], the power spectral density is used to identify possible failures of Erbium-doped fiber amplifiers (EDFAs), filters, and fibers. However, all of these algorithms require high amounts of training data to achieve high accuracies for the classification tasks. In [5], a variational autoencoder (VAE) is used for data augmentation for the input features of neural network (NN) based classifiers to reduce the overall required number of multiplications and complexity of the soft-failure identification. The input features include BER, optical signal-to-noise ratio (OSNR) and the input powers of the channels in an end-to-end monitoring scenario. However, this approach lacks the investigation of required real-world soft-failure samples. In this paper, we show the potential of data augmentation for spectral-data based soft-failure identification by comparing a VAE and a generative adversarial network (GAN). The data augmentation is based on soft-failure spectra from an experimental setup and the data is used to compare the performance of different ML-algorithms for different numbers of available real soft-failure data.

2. Dataset Generation

A dataset is generated for training and testing the framework, utilizing the experimental setup shown in Fig. 1. A bandwidth loaded WDM signal with 1, 3 and 5 channels are transmitted over three spans of 88.4 km standard single mode fiber (SSMF) with the center channel carrying a dual-polarization QAM signal. The detailed system parameters and setup are depicted in Fig. 1 [6]. The transmission spectrum is acquired with an optical spectrum analyzer (OSA) having a 10 pm resolution after the EDFA on the receiver side. The experimental setup enables the emulation of various soft-failures. To emulate an EDFA noise increase, a variable optical attenuator (VOA) is placed at the mid stage access and the attenuation levels are varied from 0.2 to 2 dB, in 0.2 dB increments, for every EDFA in the link. A laser drift is emulated by varying the transmit laser of the center channel by -2.5 to 2.5 GHz, in 0.5 GHz steps. A laser power degradation is produced by reducing the power from -2.5 to 2.5 dB in 0.5 dB increments. The same is done for the loaders by randomly selecting a channel with the same magnitude in the wavelength selective switch (WSS). Two types of filter failures are considered: filter tightening and filter shift. For filter tightening, the WSS filter characteristic is narrowed for the channels from 1 to 5 GHz in 1 GHz increments. Filter shifting is accomplished by adjusting the WSS center frequency in 1 GHz increments from -2 to 2 GHz. The various failure types were examined for the specified transmission link configurations and each type was tested twice to eliminate variations. As a result, approximately 800 spectra were recorded per failure type, resulting in a total of 5600 spectra for the failure dataset.



Experimental System Parameters	
Modulation format	DP-QPSK, DP-8-QAM, DP-16-QAM
Symbol rate	32 Gbaud
Number of channels	1, 3, 5
Channel spacing	37.5 GHz
Launch power per channel	-3, -2, -1, 0 dBm

Fig. 1. Experimental setup and data augmentation chain with a generative adversarial network (GAN) and a Variational Autoencoder (VAE) for soft-failure identification, *L*: latent space; DAC: digital-to-analog converter, ASE: amplified-spontaneous emission, WSS: wavelength selective switch, EDFA: Erbium-doped fiber amplifier, SSMF: standard single-mode fiber, OSA: optical spectrum analyzer, ADC: analog-to-digital converter, DSP: digital signal processing.

3. Data Augmentation

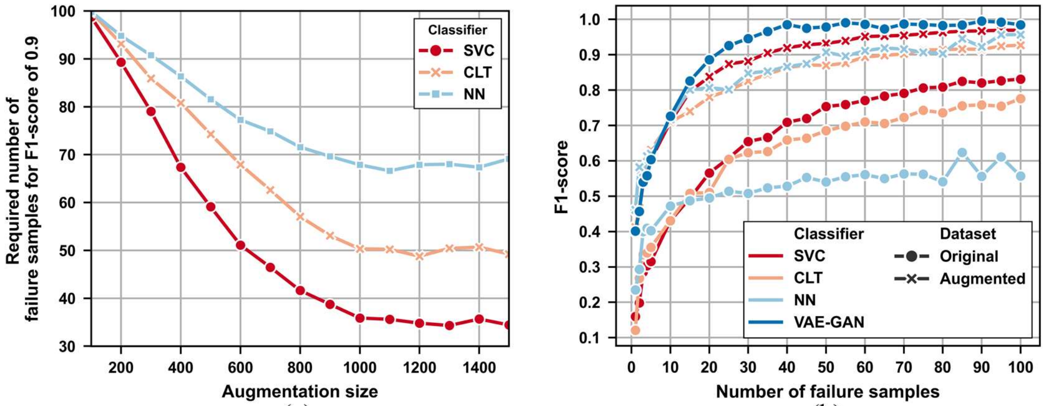
Autoencoders (AEs) have been used for various tasks like anomaly detection or input feature reduction. Conventional AEs consists of two NNs, i.e., an encoder and a decoder. The purpose of the encoder is to represent the input with a lower number of features in the so-called latent space. The decoder then reconstructs the input of the encoder from the latent space. An AE is trained to minimize the overall reconstruction error between input and output of the model. A VAE extends conventional AEs to be capable of generating new samples. The stochastic nature of the latent space gives the opportunity to generate outputs from the decoder by drawing latent space variables from its normal distribution, because the distribution of its latent space is known in contrast to conventional AEs. This leads to both encoder and decoder being probabilistic. For the case of generating new spectra, a sample is drawn from the VAE's normal distribution and transformed to a spectrum through the decoder as shown on the right side of Fig. 1. To ensure the best possible reconstruction being used for the processing in further identification algorithms, the Euclidean distance between each input and each output spectrum is calculated and the k generated spectra with the smallest Euclidean distance to the inputs are selected, where k is the desired augmentation size.

The generative capabilities of the VAE enable its usage in a generative adversarial network (GAN) for the generation of more realistic output spectra. GANs are a generative model building technique that is based on game theory [7]. The aim of a GAN is to train a generator network that can produce samples from a given data distribution by changing vectors of noise. A discriminator network is also trained to differentiate between an actual input and a generated input. As a result of this adversarial strategy, both models improve together and learn from each other. For the training of the framework, the gradient is passed through the VAE, the generator, which uses the decoder from the VAE, and the discriminator. This results in the VAE not only being trained for optimal reconstruction performance but also on separating the latent space in such a way that the discriminator can distinguish the soft-failure classes.

For evaluating the two data augmentation techniques the soft-failure identification performance is compared for various available soft-failure dataset sizes. For the VAE-based approach, the identification is done by a support vector machine-based classifier (SVC), a classification tree (CLT) and a neural network (NN) for a non-augmented case and the cases of data augmentation with the augmentation size k. It has to be noted that for the augmentation the number of failure samples is used to train the VAE and the GAN, while the testing is done on the remainder of the experimentally generated dataset. All ML-algorithms are optimized using an exhaustive grid search over 80,000 configurations. This led to the VAE encoder having an input layer of 501, a hidden layer of 25, batch normalization, and an output layer of 12. Consequently, the latent space size and input layer size of the VAE decoder are both 12. The decoder mirrors the encoder. The layers use the ReLU function as the activation function. The GAN discriminator consists of an input layer (size 501), two hidden layers (sizes 85 and 42, respectively), and an output layer equal to the number of failure classes (i.e., five).

4. Results and Discussion

The results from the investigation are summarized in Fig. 2. To show the advantage of data augmentation, the VAE and the GAN are trained on a certain number of failure samples and the F1-score of the soft-failure identification with and without usage of the augmented data is calculated. First, the optimal number of augmentation samples for each algorithm is searched using a sweep over various augmentation sizes as it is shown in Fig. 2(a). Here, the required



(a) (b)

Fig. 2. (a) Required number of real failure samples to reach an F1-score of 0.9 over the augmentation size from those samples; (b) F1-score over number of failure samples used to train the VAE and the GAN for data augmentation with an augmentation size of 1000; SVC: Support vector machine-based classifier, CLT: Classification tree, NN: Neural network

number of failure samples to reach an F1-score of 0.9 are determined for various augmentation sizes. It can be seen, that an augmentation size of 100 needs at least 98 failure samples to reach an F1-score of 0.9 across all of the three classifiers. All classifiers seem to perform best with an augmentation size over 1000. The SVC shows the best performance, followed by the CLT, and the NN performs worst. To get a better understanding of the number of failure samples needed to reach certain accuracies, the algorithms are trained on various available soft-failure training data amounts and tested on the remainder of the experimental dataset. Fig. 2(b) shows the achieved F1-score over the number of available failure samples for both the non-augmented case and a data augmentation with 1000 samples. It can be seen that overall the classifiers without using augmented data show a poor performance. The SVC reaches an F1-score of slightly above 0.8 over 80 failure samples, while the CLT reaches an F1-score of 0.7521 at the same number of samples. The neural network seems to require a lot of training data resulting in low F1-scores across all given number of failure samples for training reaching a maximum of 0.6213 at 85 training samples. The behavior changes when the VAE-based data augmentation is used. Here all of the three classifiers reach F1-scores above 0.8 between 15 and 25 available failure samples before eventually surpassing 0.9 in all cases. The SVC tops out at 0.9726 whereas the CLT and NN both show a maximum at around 0.94. Above all of those algorithms is the VAE-based GAN, which reaches F1-scores above 0.9 at a total of 25 failure samples being available for training. Those 25 samples correspond to 5 failure samples per failure class. The GAN reaches a maximum F1-score of 0.9873 at only 40 failure samples.

5. Conclusion

We show that the usage of data augmentation methods reduces the amount of needed real-world soft-failure data in order to perform well in identification tasks. The performance of an SVC, a CLT and an NN are compared based on the number of available soft-failure training samples in a non-augmented and augmented scenario. It is shown, that data augmentation is advantageous in such missing data scenarios reaching F1-scores above 0.9 with only 35 available training samples. Furthermore, we propose the usage of a VAE-GAN as an advanced machine learning algorithm to perform the soft-failure identification in order to have an even higher tolerance for missing data. The GAN shows the highest F1-score at low amounts of available training data, surpassing 0.9 at only 25 soft-failure training samples, which corresponds to 5 failure spectra per failure type.

Acknowledgements

This work has been performed in the framework of the CELTIC-NEXT project AI-NET-PROTECT (Project ID C2019/3-4), and it is partly funded by the German Federal Ministry of Education and Research (16KIS1284).

References

[1] Rafique, et al., “Cognitive assurance architecture for optical network fault management.”, in JLT, vol. 36.7 (2017): 1443-1450.
[2] Musumeci, et al., “A tutorial on machine learning for failure management in optical networks.”, in JLT, vol. 37.16 (2019): 4125-4139.
[3] Vela, et al., “BER degradation detection and failure identification in elastic optical networks.”, in JLT, vol. 35.21 (2017): 4595-4604.
[4] Lun, et al., “Soft Failure Identification for Long-Haul Optical Communication Systems Based on One-Dimensional Neural Network by Exploiting Digital Spectrum Information.”, in JLT, vol. 38.11(2020):2992-2999.
[5] Khan, et al., “Data augmentation to improve performance of neural networks for failure management in optical networks.”, in JOCN, vol. 15.1 (2023):57-67.
[6] Kruse, et al., “Experimental Investigation of Spectral Data Enhanced QoT Estimation.”, JLT, vol. 41.18 (2023): 5585-5894.
[7] Goodfellow, et al., “Generative adversarial networks.”, Communications of the ACM, vol. 63.11 (2020):139-144.