Accurate EDFA Gain Modelling using Convolutional Neural Network with Denoising Layers enabled by Soft-thresholding

Hongbo Chen(1), Guo-Wei Lu(1) (*), Kunping Luo(1)

(1) Division of Computer Engineering, The University of Aizu, Fukushima 965-8580, Japan
(*) gordon.guoweilu@gmail.com

Abstract *We propose an end-to-end EDFA gain model using denoising CNN. Our proposed model achieves state-of-the-art performance on both the same device used for training and different physical units of the same make, with the lowest RMSEs compared to conventional CNN and previously-reported NN schemes. ©2023 The Author(s).*

Introduction

The ever-increasing demand for higher data transmission rates in contemporary telecommunications infrastructure has put considerable pressure on long-distance and regional fibre optic networks. In order to counteract signal attenuation during transmission, erbium-doped fibre amplifiers (EDFAs) are extensively employed to boost optical signals. Consequently, precise modelling of the EDFA gain profile becomes crucial for effective network topology management and capacity optimisation [1]. Inaccurate models may result in substantial discrepancies between the predicted signal power and the actual power, which can potentially cause error accumulations in optical transmission links containing multiple EDFAs. Therefore, it is of paramount importance to develop a highly accurate EDFA gain model.

Neural networks (NN), specifically multi-layer perceptron (MLP), have been extensively utilized for EDFA gain modelling in various studies [2-5]. Previous research [2-3] has predominantly concentrated on training and testing using the same EDFA device. However, employing the gain model derived from one device on others may lead to diminished accuracy owing to differences in device characteristics. Consequently, enhancing the generalization capability of machine learning-based EDFA gain models is of vital importance in order to tackle this issue.

In [4], a differentiable neural network model was introduced to achieve more precise modelling by employing three times the typical amount of training data. This model demonstrated a reduction in mean square error (MSE) from approximately 0.06 to 0.02 dB², corresponding to a decrease in root mean square error (RMSE) from 0.25 dB to 0.14 dB. Furthermore, in [5], an auxiliary neural network was proposed to diminish discrepancies among different EDFAs by conducting retraining with a limited amount of data. Nonetheless, both approaches necessitate additional training data or a retraining process, which can prove to be computationally demanding and may not be viable in certain situations.

Tab. 1: Performance Comparison of EDFA Gain Models

Model	Portability	E2E	No Extra Data
NN [2]	×	√	×
Multiple NNs [3]	×	√	×
Differentiable NN [4]	√	√	×
Auxiliary NN [5]	√	×	×
Ours	√	√	√

In this work, we put forth an end-to-end EDFA gain model that capitalizes on a denoising convolutional neural network (CNN) with remarkable generalization capabilities, without necessitating additional training data or re-training. Our method integrates soft-thresholding-based denoising transform layers into the CNN architecture to enhance both the modelling accuracy and generalization capacity. For the same device used for training, our model achieves an impressive RMSE of a mere 0.05 dB. Moreover, due to the robustness reinforced by the proposed denoising layers and the local bias property of CNN, when transferring the trained model to different physical units of the same make, the RMSE remains low at 0.05 dB for low gain ranges (<15 dB) at 15 dBm output power. These results attest to the robustness and superiority of our approach in terms of accuracy and generalization, as shown in Table 1.

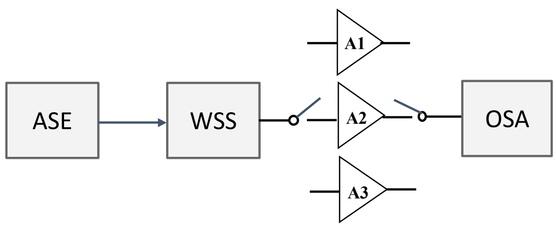

Fig. 1: Experimental setup for collecting datasets. ASE: Amplified Spontaneous Emission, WSS: Wavelength Selective Switch, Optical Spectrum Analyzer.

Denoising enabled by Soft Thresholding

As a classical signal denoising technique, soft thresholding [7] has been extensively employed in signal processing over the past decade to mitigate noise or suppress minor signal scale. By integrating soft thresholding as denoising non-linear transformation layers into the CNN, the learning capability of the neural network can be substantially enhanced when dealing with highly-noisy datasets. Consequently, this improves the accuracy of the model for CNN-based EDFA gain modelling. Soft thresholding function can be expressed by the following equation [8]:

$$y = \begin{cases} x-\tau, & x>\tau \\ 0, & -\tau \le x \le \tau \\ x+\tau, & x<-\tau \end{cases} \qquad (1)$$

where x presents the input feature, y is the output, and τ is the threshold. Soft thresholding maintains the valuable negative or positive features whilst suppressing noise by setting near-zero features to zeros, thereby offering excellent denoising capability. To ensure the efficacy of the soft threshold's denoising ability, it is essential to estimate the threshold appropriately. To achieve this, a self-attention mechanism [8-9] is incorporated to learn an adaptive threshold. As depicted in Fig. 2, a threshold unit comprising a global average pooling (GAP), a fully connected (FC) layer, and a sigmoid activation function (Sigmoid) is employed to implement an attention operation and compute an adaptive threshold (τ). It is worth noting that, in order to preserve the negative features, the absolute value is utilised in threshold learning. Following the soft thresholding operation, the resulting features are multiplied by the signum function output of the original input features. It is postulated that soft thresholding could effectively enhance the model's accuracy by suppressing noise in the input features.

Dataset and Data Pre-processing

To assess the performance of the proposed model, we conduct experiments using a publicly-available dataset [6]. Fig. 1 illustrates the experimental setup for gathering the datasets utilized in training and testing. A wavelength selective switch (WSS) shapes the spectrum of a fattened amplified spontaneous emission (ASE) source, ranging from 191.5 THz to 196.25 THz. A1, A2, and A3 denote three distinct booster EDFAs of the same manufacturer, respectively. One of the three EDFAs (A1, A2, and A3) is fed with the shaped spectrum as input, and the resulting optical spectra are measured through an optical spectrum analyser (OSA). The datasets contain input and output power spectral

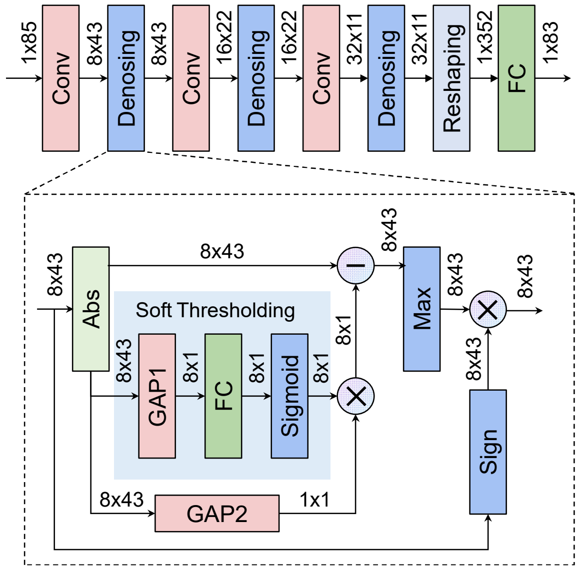

Fig. 2: The structure of the proposed EDFA gain model based on CNN with inserted denoising layers.

density (PSD) profiles, as well as total input (P_{in}) and output power (P_{out}) pairs. The input PSD contains 83 equally-spaced frequency channels in the C-band. For each EDFA, the spectra with four distinct output powers (15 dBm, 16 dBm, 17 dBm, and 18 dBm) were measured and included in the datasets. To ensure a fair comparison, the same pre-processing method as that in [5] is applied to the datasets.

EDFA Model

The proposed model architecture is shown in Fig. 2, which comprises three convolutional layers, three denoising layers, and one FC layer for processing extracted features. The primary objective of this model is to accurately extract and represent features from the input data while effectively reducing noise and achieving more precise gain modelling. Each 1-dimensional convolutional layer is followed by a batch normalization layer designed to learn spatial patterns from the input data. The kernel size of the convolution is 3, and the stride is 2. The corresponding sizes of the feature maps for each layer are also illustrated in Fig. 2. As discussed in the previous section, the proposed soft-thresholding can effectively improve the model's accuracy by reducing the influence of noise on the input data. Finally, the FC layer performs regression tasks based on the features extracted from the input data, which takes the output of the convolution and denoising layers and applies a set of weights to produce the final result.

Performance Investigation

The performance of the proposed model is evaluated by quantifying the predictive accuracy, which is measured by calculating the RMSE between the predicted output spectrum and the corresponding measured spectrum. Two experiments are conducted: intra-EDFA, where training and testing are performed on EDFA-A1, and inter-EDFA, where training is conducted on EDFA-A1 and testing is carried out on EDFA-A2 and EDFA-A3. For performance comparison, three models are employed in each experiment: the NN proposed in [5], the proposed denoising CNN model, and the conventional CNN without denoising layers, referred to as CNN here. We apply the same pre-processing method to all experiments to ensure a fair comparison.

The intra-EDFA results are shown in Fig. 3(a)-(d). To ensure the validity of our experiment, a cross-validation approach is adopted. Specifically, the training dataset is evenly partitioned into five portions, corresponding to the five folds on the horizontal axis. Four folds are used for training, and one is reserved for testing. The proposed model consistently exhibits significantly lower error than the other two models, with a RMSE of approximately 0.05 dB. Furthermore, the proposed model demonstrates stable performance in cross-validation when operating at different power levels. Importantly, compared to the CNN model, the proposed CNN model with denoising layers significantly improves the modelling accuracy, verifying the effectiveness of denoising enabled by soft thresholding.

The inter-EDFA performance is presented in Fig. 3(e)-(h). Unlike the intra-EDFA experiment, the RMSE performance of the inter-EDFA experiment exhibits slight dependence on the EDFA gain. Nonetheless, the proposed model outperforms the other two models for four different EDFA output powers. Furthermore, the model maintains comparable RMSE when transferred to different physical units. For instance, when the output power is 15 dBm, the RMSE of the inter-EDFA experiment in the low gain regime (Gain<15 dB) depicted in Fig. 3(e) is comparable to that observed in the intra-EDFA experiment shown in Fig. 3(a). This outcome indicates the high generalizability of the proposed denoising CNN model.

Conclusions

A precise, highly-generalizable, robust EDFA gain model is proposed based on convolutional neural network with denoising layers enabled by soft thresholding. Compared to previous work, this model demonstrates scalability in end-to-end training and does not rely on additional data. This precise and scalable EDFA model paves the way for broader application explorations.

Acknowledgements

This work was partially supported by the JSPS KAKENHI grant (22H01490) and the NICT Project grant (02101).

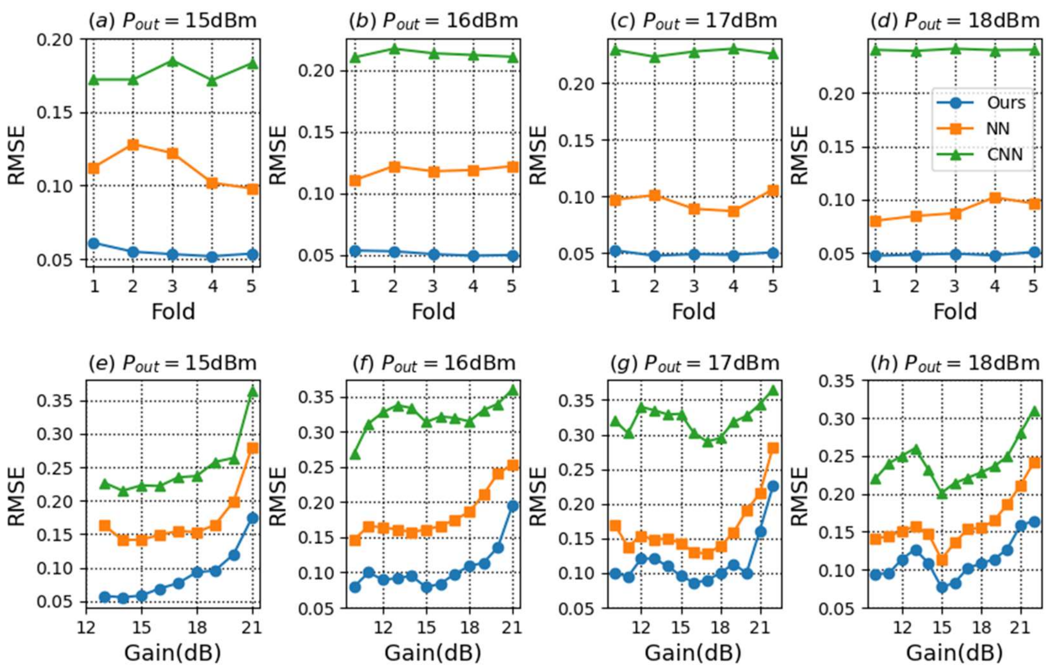

Fig. 3: (a)-(d): Intra-EDFA RMSE for different folds of data with output powers of (a) 15dBm (b) 16dBm (c) 17dBm (d) 18dBm; **(e)-(h):** Inter-RMSE vs. Gain with power powers of (e) 15dBm (f) 16dBm (g) 17dBm (h) 18dBm.

References

[1] Rafique, Danish, and Luis Velasco. "Machine learning for network automation: overview, architecture, and applications." *Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D126-D143, Oct. 2018, doi:10.1364/JOCN.417434

[2] Y. You, Z. Jiang and C. Janz, "Machine Learning-Based EDFA Gain Model," *European Conference on Optical Communication (ECOC)*, Rome, Italy, 2018, pp. 1-3, doi: 10.1109/ECOC.2018.8535397.

[3] S. Zhu, C. L. Gutterman, W. Mo, Y. Li, G. Zussman and D. C. Kilper, "Machine Learning Based Prediction of Erbium-Doped Fiber WDM Line Amplifier Gain Spectra," *European Conference on Optical Communication (ECOC)*, Rome, Italy, 2018, pp. 1-3, doi: 10.1109/ECOC.2018.8535323.

[4] F. da Ros, U. C. de Moura and M. P. Yankov, "Machine learning-based EDFA Gain Model Generalizable to Multiple Physical Devices," *European Conference on Optical Communications (ECOC)*, Brussels, Belgium, 2020, pp. 1-4, doi: 10.1109/ECOC48923.2020.9333297.

[5] J. Lin, X Lin and Z.Jiang, "Auxiliary Neural Network Assisted Machine Learning EDFA Gain Model," *Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, USA, 2023, M2E.2

[6] Yankov, Metodi Plamenov; Da Ros, Francesco, "Input-output power spectral densities for three C-band EDFAs and four multi-span inline EDFAd fiber optic systems of different lengths," *Technical University of Denmark. Dataset*, 2020, doi: 10.11583/DTU.13135754.v1

[7] D. L. Donoho, "De-noising by soft-thresholding," in *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613-627, May 1995, doi: 10.1109/18.382009.

[8] M. Zhao, S. Zhong, X. Fu, B. Tang and M. Pecht, "Deep Residual Shrinkage Networks for Fault Diagnosis," in *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4681-4690, July 2020, doi: 10.1109/ TII. 2019. 2943898.

[9] J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu, "Squeeze-and-Excitation Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 8, pp. 2011-2023, 1 Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.