Self-Normalizing Neural Network, Enabling One Shot Transfer Learning for Modeling EDFA Wavelength Dependent Gain

Agastya Raj*,(1), Zehao Wang(2), Frank Slyne(1), Tingjun Chen(2), Dan Kilper (1), Marco Ruffini(1)

(1) CONNECT Centre, School of Computer Science and Statistics and School of Engineering, Trinity College Dublin, Ireland, *rajag@tcd.ie
(2) Duke University, Department of Electrical and Computer Engineering, Durham, NC, USA

Abstract *We present a novel ML framework for modeling the wavelength-dependent gain of multiple EDFAs, based on semi-supervised, self-normalizing neural networks, enabling one-shot transfer learning. Our experiments on 22 EDFAs in Open Ireland and COSMOS testbeds show high-accuracy transfer-learning even when operated across different amplifier types.* ©2023 The Author(s)

Introduction

The gain spectrum of an Erbium-Doped Fiber Amplifier (EDFA) has a complex dependence on channel loading, pump power, and operating mode, making accurate modeling difficult to achieve. Recently, Machine Learning (ML) techniques such as Neural Networks (NNs) have been used to build EDFA gain models[1],[2]. Other work[3] has produced generalized ML-based EDFA models using training datasets collected from multiple EDFAs of the same make and model, which are shown to achieve lower Mean Absolute Error (MAE) of the gain spectrum prediction across multiple devices of the same make. Although these models achieve high prediction accuracy, they do require a large number of measurements, which can be time-consuming and difficult to obtain if the EDFA is in a live network. Due to the complexity of the model, NN also suffer from non-convex training criteria and local minima, which complicate the training process especially with limited number of measurements.

Transfer Learning (TL) techniques[4] have been recently used to try and mitigate this issue, by training a base model on one EDFA and then using this to model different devices, by only using minimal additional data from the new device. Recently, it was demonstrated[5] that a single EDFA model can be transferred between different EDFAs of the same type using only 0.5% of the entire dataset, showcasing the potential for efficient model transfer in this domain. Yet, the application of transfer learning across amplifiers of different types (i.e., from a EDFA Booster base model towards an EDFA Preamp target model) requires further investigation. In addition, work to date has mostly relied on training data from external features, such as input power levels and output gain spectra, which may not fully capture the complex behaviour of EDFAs.

In this paper, we implement and study a novel semi-supervised, self-normalizing NN approach (hereafter referred to as the SS-NN model) that characterizes the wavelength-dependent gain of an EDFA using just 256 labeled measurements along with additional unlabeled data (which are easier to obtain). By incorporating internal EDFA features that are typically available in commercial telecom equipment, our model can be transferred to different EDFA types with only a single new measurement through transfer learning. We evaluate our approach on 22 different EDFAs across the Open Ireland (based in Dublin, Ireland) and PAWR COSMOS (based in Manhattan, USA) testbeds, achieving a MAE within 0.14 dB for same-type transfers and 0.17 dB for cross-type transfers.

EDFA Gain Spectrum Measurement Dataset

We carry out gain measurements across multiple wavelengths in the C-band from 3 commercial-grade Lumentum ROADM-20 units deployed in the Open Ireland testbed[6] and 8 similar units deployed in the PAWR COSMOS testbed[7],[8], each with 2 EDFAs (a Booster and a Pre-Amplifier), resulting in a total number of 22 EDFAs. To ensure consistency, we followed a similar measurement setup and data collection pipeline for both testbeds[9]. In the Open Ireland testbed, all EDFAs were measured at target gains of 15/20/25 dB, while in the COSMOS testbed, the target gains were 15/18/21 dB for Boosters and 15/18/21/24/27 dB for Pre-Amplifiers in high gain mode with 0 dB gain tilt (we adopt different gain setting to emulate diversity of operation in different networks). The dataset includes 3,168 gain measurements (at multiple wavelengths) for each EDFA, for each given target gain settings, across 95×50 GHz channels in the C-band. In addition, measurements for each EDFAs are collected under two channel loading modes: Random and Goalpost allocation (i.e., loading groups of chan-

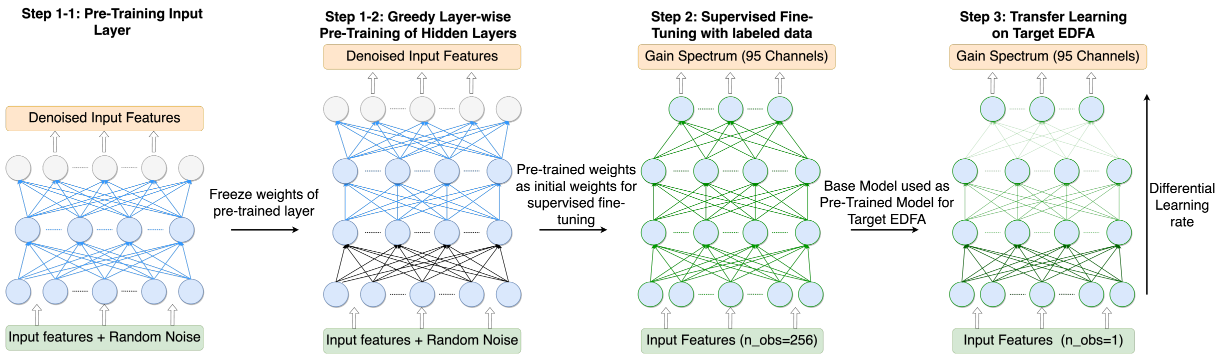

Fig. 1: Model Training and Transfer Learning (TL) Framework

nels in different spectrum bands).

Training of the Base Model

We construct a base model (used as the *source model* in the transfer learning process) for both booster and pre-amplifier EDFAs using a 5-layer NN with neurons initialized with zero weights. The NN architecture consists of 200 neurons in the first two layers, 100 neurons in the next two layers, and 95 neurons in the final layer, predicting the wavelength-dependent gain output. Input features to the model include EDFA target gain setting, total input/output power, input power for each channel, a binary vector for each channel representing the channel loading configuration, and three internal features related to the embedded Variable Optical Attenuator (VOA): total VOA input/output power and attenuation.

Due to the limitations of batch normalization when fine-tuning models with less than 32 observations[10], we utilize Self Normalizing Neural Networks (SNN) with Scaled Exponential Linear Unit (SELU) activation function[11] instead of the Rectified Linear Unit (ReLU). This choice enables us to effectively normalize the hidden layer outputs with a small amount of data, while maintaining the benefits of hidden layer normalization and preserving high accuracy. This step is the key enabling factor of our developed NN architecture to achieve effective one-shot training and transferability between models.

We employ a 2-step process to train the source model, including unsupervised pre-training[12][13] and supervised fine-tuning[14] (see Fig. 1):

First, in the *unsupervised pre-training* step, we initialize the source model's weights using unlabeled data from 512 measurements for each target EDFA gain setting. Noise is added to the data, and the input layer is trained as an autoencoder to denoise and reconstruct the input. We construct this autoencoder by removing subsequent layers and adding a decoder layer on top. The autoencoder is trained for 1,800 epochs with a learning rate of 0.001, using the Adam Optimizer and Mean Squared Error (MSE) loss function. The weights of this layer are fixed and used as the basis for training the subsequent layers.

Second, in the *supervised fine-tuning* step, we utilize 256 randomly loaded gain spectrum measurements to train the model in a supervised manner. The model is trained using the MSE loss function across all loaded channels, with a learning rate of 0.001 over 1,200 epochs. The test set comprises all Fixed Goalpost (270 measurements) and 20% of the Random Baseline (220 measurements) EDFA gain spectrum measurements, so that we can compare the performance for the two different channel loading scenarios[5].

Transfer Learning (TL) to Target EDFA

To transfer from a source model to a target EDFA, we re-train the same model using a single randomly loaded measurement for 10,000 epochs, using MSE as the loss function and Adam optimizer. We use a differential learning rate across layers, where the output layer has a larger learning rate of 1e-03 compared to the subsequent hidden layers, which have progressively decreasing learning rates, with each layer's rate being 10% of the next layer's rate. In this way, the weights of the output layer are modified more aggressively, allowing it to capture the specific characteristics of the target EDFA more effectively. At the same time, the lower levels of the NN are fine-tuned more gradually to avoid overfitting and ensure that the model can be generalized to new inputs.

Results

We compare our SS-NN based TL technique with a benchmark state-of-the-art method[5], using the same set of features to highlight the benefit of our approach. Additionally, we demonstrate the advantage of incorporating internal EDFA features by comparing the results of the SS-NN model with and without these additional features.

TL to the Same EDFA Type: Figs. 2(a) and 2(b) present the MAE of the three approaches for TL for source Booster→Target Booster and Source

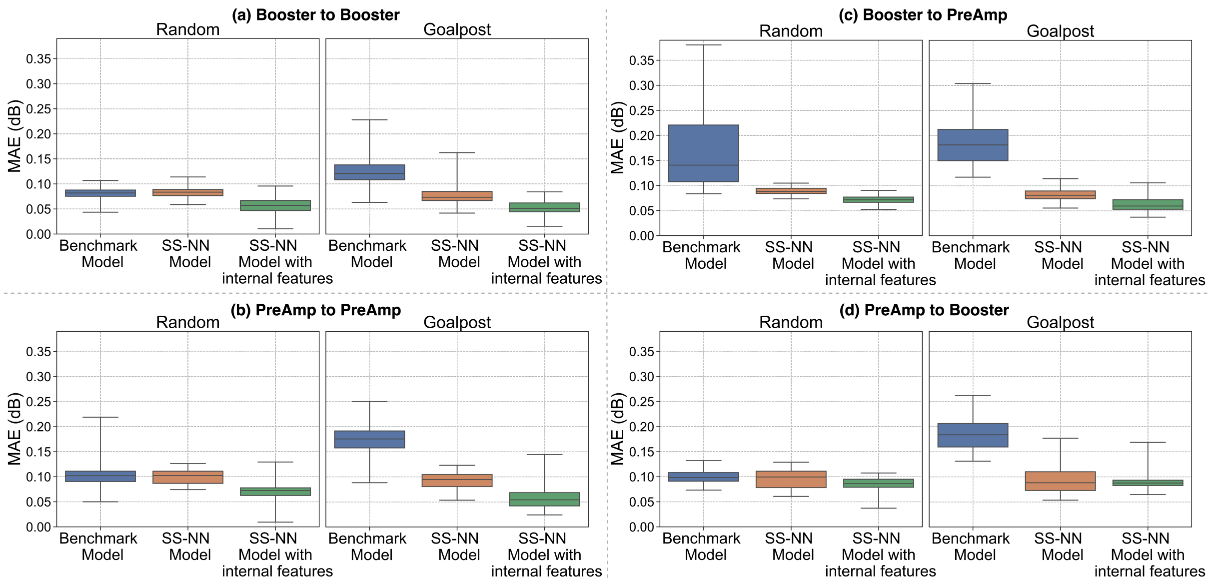

Fig. 2: Boxplot distribution of MAE across 22 EDFAs of (a) Booster to Booster TL, (b) PreAmp to Preamp TL, (c) Booster to Preamp TL and (d) Preamp to Booster TL. The boxes denote the inter-quartile range, and the whiskers denote the min/max.

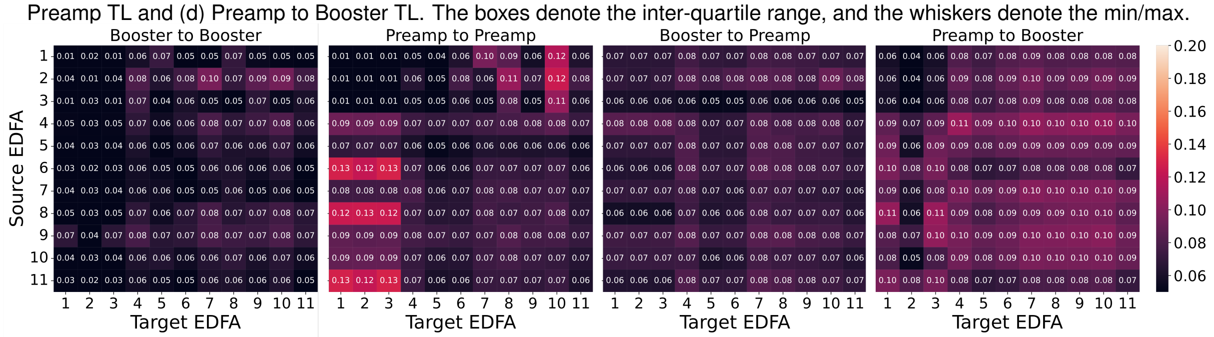

Fig. 3: Transfer Learning MAE matrix of SS-NN model with internal features on random loading. The (i, j) entry corresponds to the TL-based EDFA model, where the i^{th} and j^{th} EDFA serve as the source and target models, respectively. EDFA 1-3 are deployed in Open Ireland, while EDFA 4-11 are deployed in COSMOS.

Pre-Amplifier→Target Pre-Amplifier respectively, for both random and goalpost channel loading, across 22 EDFAs. Our SS-NN model outperforms the benchmark technique for goalpost loading and exhibits comparable performance for random channel loading. In addition, when we include additional features, we see improvement in both channel loading configurations.

TL to Cross-EDFA Types: Figs. 2(c) and 2(d) report the MAE of the three approaches for TL between different types of EDFAs (Booster→Preamp and Preamp→Booster) respectively, under random and goalpost channel loading configurations. The SS-NN model again displays significant improvement and consistency in cross-type transfer, with a 3× improvement in MAE over the benchmark algorithm even if using the same features. Including internal VOA features further improves performance, leading to similar performance as the same-type transfer.

A key advantage of the SS-NN model is its consistent performance across TL between all EDFAs. Fig. 3 shows the MAE of the TL model incorporating internal features on random channel loading. The model demonstrates consistent performance in terms of the EDFA gain spectrum prediction accuracy, with an MAE within 0.13 dB for same-type transfers and within 0.11 dB for cross-type transfers.

Conclusions

We analyze a novel semi-supervised learning technique to model the gain spectrum of an EDFA using a minimal amount of data. The model can be transferred to EDFAs of different types using a single new measurement, showing that a single EDFA can be used to characterize multiple EDFAs using minimal data collection. We also find that using internal EDFA features available to the operator provides enhanced performance in both same-type and cross-type transfers, showing potential for improvement by incorporating internal features.

Acknowledgements

Supported by grants from SFI: 12/RC/2276_p2, 18/RI/5721, and 13/RC/2077_p2 and NSF: CNS-1827923, OAC-2029295, and CNS-2112562.

References

[1] 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”, in *2018 European Conference on Optical Communication (ECOC)*, Rome: IEEE, Sep. 2018, pp. 1–3, ISBN: 978-1-5386-4862-9. DOI: 10.1109/ECOC.2018.8535323. (visited on 10/10/2022).

[2] S. Zhu, C. Gutterman, A. D. Montiel, *et al.*, “Hybrid Machine Learning EDFA Model”, in *Optical Fiber Communication Conference (OFC) 2020*, San Diego, California: Optica Publishing Group, 2020, T4B.4, ISBN: 978-1-943580-71-2. DOI: 10.1364/OFC.2020.T4B.4. (visited on 10/10/2022).

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

[4] F. Zhuang, Z. Qi, K. Duan, *et al.*, “A Comprehensive Survey on Transfer Learning”, *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, ISSN: 1558-2256. DOI: 10.1109/JPROC.2020.3004555.

[5] Z. Wang, D. Kilper, and T. Chen, “Transfer learning-based ROADM EDFA wavelength dependent gain prediction using minimized data collection”, in *Optical Fiber Communication Conference (OFC) 2023*, ser. Optical Fiber Communication Conference (OFC) 2023, Optica Publishing Group, 2023, Th2A.1.

[6] CONNECT, “Openireland testbed, funded by science foundation ireland”, 2022. [Online]. Available: www.openireland.eu.

[7] T. Chen, J. Yu, A. Minakhmetov, *et al.*, “A Software-Defined Programmable Testbed for Beyond 5G Optical-Wireless Experimentation at City-Scale”, *IEEE Network*, vol. 36, no. 2, pp. 90–99, Mar. 2022, ISSN: 1558-156X. DOI: 10.1109/MNET.006.2100605.

[8] D. Raychaudhuri, I. Seskar, G. Zussman, *et al.*, “Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless”, in *Proc. ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2020.

[9] Z. Wang, D. Kilper, and T. Chen, *An Open EDFA Gain Spectrum Dataset and Its Applications in Data-driven EDFA Gain Modeling*, Apr. 2023. DOI: 10.1364/opticaopen.22661170.v1. (visited on 04/26/2023).

[10] S. Ioffe, “Batch renormalization: Towards reducing minibatch dependence in batch-normalized models”, in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, *et al.*, Eds., vol. 30, Curran Associates, Inc., 2017.

[11] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, *Self-Normalizing Neural Networks*, Sep. 2017. DOI: 10.48550/arXiv.1706.02515. arXiv: 1706.02515 [cs, stat]. (visited on 04/26/2023).

[12] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layer-wise training of deep networks”, in *Advances in Neural Information Processing Systems*, B. Schölkopf, J. Platt, and T. Hoffman, Eds., vol. 19, MIT Press, 2006.

[13] J. Ge, S. Tang, J. Fan, and C. Jin, *On the Provable Advantage of Unsupervised Pretraining*, Mar. 2023. DOI: 10.48550/arXiv.2303.01566. arXiv: 2303.01566 [cs, math, stat]. (visited on 04/26/2023).

[14] N. J. Prottasha, A. A. Sami, M. Kowsher, *et al.*, “Transfer learning for sentiment analysis using BERT based supervised fine-tuning”, *Sensors*, vol. 22, no. 11, 2022, ISSN: 1424-8220. DOI: 10.3390/s22114157.