Iterative Transfer Learning Approach for QoT Prediction of Lightpath in Optical Networks

Iterative Transfer Learning Approach for QoT Prediction of Lightpath in Optical Networks

Hafsa Tariq[1], **Fehmida Usmani**[1], **Ihtesham Khan**[2], **Muhammad Umar Masood**[2], **Arsalan Ahmad**[1] , **Vittorio Curri**[2]

[1]*National University of Sciences & Technology (NUST), Islamabad, Pakistan*

[2]*Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129, Torino, Italy*

e-mail: fusmani.dphd18seecs@seecs.edu.pk

ABSTRACT Machine learning (ML) has been widely used in optical networks for accurate Quality-of-transmission (QoT) estimation of Lightpaths (LPs). However, this domain has two main issues: ML-based models require a sufficiently large amount of data for training, and once the model is trained on one type of configuration, it cannot be used for another configuration. This paper focuses on these two issues and proposes an Active Transfer Learning (ATL) based solution. In ATL, Active learning (AL) helps in reducing the dataset's size while not compromising the model's performance, while the Transfer learning (TL) concept enables the transfer of knowledge from a source domain to the target domain with improved accuracy. This combined approach of ATL delivers promising results with minimum data samples and enhanced performance.

Keywords: Active learning, Transfer learning, Quality-of-transmission, Deep neural network

1. INTRODUCTION

In optical networks, sensing whether an LP is adequate for transmission before its deployment in the network is an efficient way to conserve resources. Nowadays, ML has been a center of attention for the QoT prediction of the LP [1]–[3], but this domain requires a sufficiently large amount of data for training and testing and is not suitable enough if the train and test datasets hold different feature spaces or distributions. As most of the optical networks are established with varying configurations, an ML model trained on one type of configuration does not perform well for another configuration, and practically the optical networks have a limited number of active LPs in them which means that fewer labeled samples are available for training [4]. This paper addresses these two issues and proposes an ATL approach, which combines the benefits of both TL and AL approaches. TL deals with the first problem and applies the concept that a model trained on one type of configuration can be used to train the model for another related problem by transferring the learned knowledge. TL is proved to be a very efficient approach as it enables the target domain model to provide maximum accuracy even with a limited number of samples for training, while AL addresses the second problem and improves the model's performance while using only a sufficient amount of labeled samples and helps in reducing the labeling cost [5], [6]. AL enables the model to select the most informative data samples and achieve higher accuracy with minimal labeled training samples. An AL approach selectively requests the oracle to provide the labels for the most effective data samples and increases its training dataset after every iteration, while minimizing the labeling cost for the samples. In this work, a novel ATL-based framework is proposed, where a model trained with the AL approach is used for knowledge transfer to another model to enable adaptability for other similar problems while using a minimum number of training samples. The target domain model is also trained with the AL approach. In the proposed framework, AL is applied to refine the Deep neural network (DNN) tuning, and TL helps to efficiently map the knowledge gained from the source domain to any related target domain. In this work, we validated the proposed ATL approach for the classification of LP based on its QoT into good or bad LP before its deployment.

2. SIMULATION SETUP AND DATASET GENERATION

A framework of software-defined optical networks is considered with an Optical line system (OLS) as the network edge and Reconfigurable optical add-drop multiplexers (ROADM) as network nodes. The amplifier's ripple gain describes the physical layer's unsettling nature while OLS is operated optimally. At the transmission layer, LPs are installed on the Wavelength division multiplexing (WDM) flexible grid to enable dual-polarization multilevel modulation format and bind the transceivers. A 50 GHz grid size with 76 C-band channels is assumed for the simulation scenario. Due to limited computational resources, only 76 channels with a total bandwidth of about 4 THz are analyzed. The transmitter produces 32 GBaud signals that are shaped by a root-raised-cosine filter. Erbium-doped fiber amplifier (EDFA), operating in a constant output power mode of 0 dBm per channel, maintains the launch power of the signal at 0 dBm. The EDFA noise figure is constant between 3.5 dB and 4.5 dB, and the ripple gain varies by only 1 dB. All links are assumed to use Standard single-mode fiber (SSMF). Fiber impairments, such as attenuation (α) = 0.2dB/km and dispersion (D) = 16ps/nm/km, are also considered. Amplified spontaneous noise (ASE) and Non-linear interference (NLI) are the most significant impairments that degrade the LP during the transmission of signals [9]. The Generalized SNR (GSNR) measurement, which contains both the aggregated effect of ASE noise and NLI disturbance, provides the QoT metric for a specific

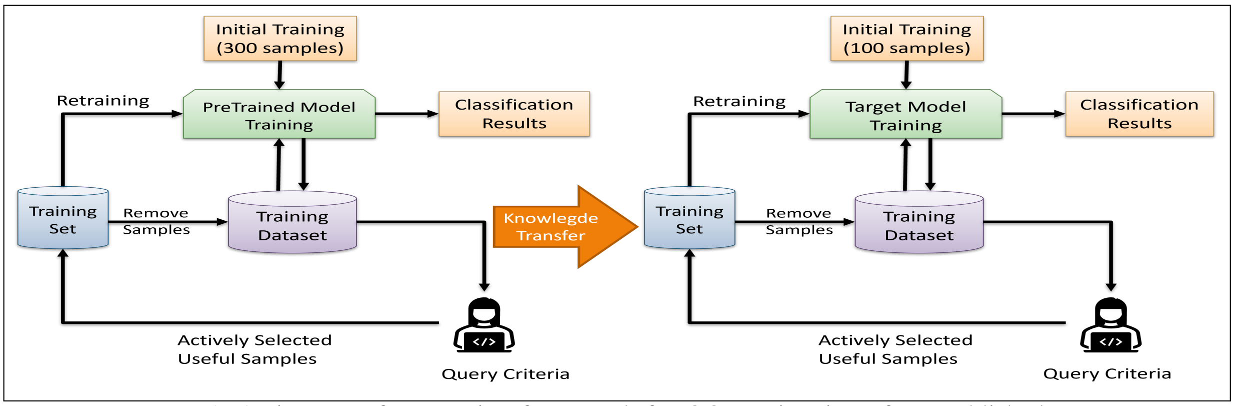

Figure 1: Active Transfer Learning framework for QOT estimation of unestablished LP.

TABLE I: Network simulation parameters.

Simulation Parameters	
Launch Power/ Channel	0 dBm
Dispersion (D)	16.0 $ps/nm/km$
Attenuation coefficient (α)	0.2 dB/km
Channel Spacing (C-Band)	50 GHz
Span Length	80 km
WDM Comb (C-Band)	76
Baud Rate (C-Band)	32 $Gbaud$
Amplifier Noise Figure	[3.5 - 4.5] dB [7]
Amplifier Gain Ripple	Variation of 1 dB
Fiber Type	Standard SMF

TABLE II: Topology details

Topology Details [8]		
Parameters	**EU: Training**	**US : Testing**
Number of Nodes	28	100
Number of Links	41	171
Average path distance (km)	2014.06	2541.75
Maximum path distance (km)	3051.10	5481.07
Minimum path distance (km)	669.30	568.33
Average number of spans/Link	19.75	27.49

LP routed via definitive OLSs from the source node to the destination node. The total $GSNR$ of each LP propagating across the OLS can be calculated using $\frac{1}{GSNR} = \sum_n \frac{1}{GSNR_n}$, where n is the number of OLSs, the LP passed along a specific path. The ASE and NLI over the particular path are both taken into account by the GSNR metric. In Table I, the specifications of the network simulation parameters are reported. To build synthetic datasets similar to the one described, we abstracted the physical layer using the open-source GNPy library [10]. The GNPy library utilizes an end-to-end simulated environment to construct physical layer network models. The synthetic dataset is generated against two distinct network topologies: the European Union (EU) network and the United States (US) network. In the given scenario, we consider the EU Network a well-deployed network representing the source domain, whereas the US Network represents the target domain. The two networks being compared are identical in terms of fiber and optical network elements. However, they differ in terms of the sensitive parameters of the amplifier (noise figure and ripple gain) and fiber insertion losses. The realization of spectral load for every simulated link in a dataset is a subset of 2^{76}, where 76 indicates the total number of channels. We analyzed 3000 realizations of arbitrary traffic flows ranging from 34% to 100% of the total operating bandwidth for each source-to-destination (s $\rightarrow$ d). The data set utilized in this study is made up of 5 pairs (s $\rightarrow$ d) pairs from the EU network and 2 pairs (s $\rightarrow$ d) from the US network. Thus, 15,000 realizations are generated for the EU network topology, whereas 6,000 are generated for the US network topology. The topology details of both the EU and the US networks are given in the Tab. II

3. Iterative transfer learning approach for LP establishment

The proposed methodology is depicted in Fig. 1. In this ATL approach, a DNN model is trained on an EU dataset. Typically, a data-driven scheme like DNN requires a large number of labeled data samples to perform effectively. In contrast, in the proposed scenario, we consider that only a few labeled samples are available, and training a model directly on these small-scale samples can lead to overfitting, which results in ineffective prediction. To avoid this issue, the proposed scheme utilized the AL approach, in which only a few high-quality data samples are selected from the samples pool to train the model with less yet informative samples. On top of this, it is combined with the TL scheme, which enables the transfer of knowledge from one domain to another. The TL approach exploits the EU network as a source domain and the US network as the target domain.

The proposed pre-trained AL-based DNN model is depicted in Fig. 2. It contains multiple inner layers with varying neurons, but all have Rectified linear unit (ReLU) activation functions to prevent the problem of vanishing gradient. The last layer has a sigmoid as an activation function for classification purposes. After each inner layer, there exists a dropout layer with a rate of 0.20 to drop 20% neurons during each iteration to prevent overfitting. This base model is trained using the AL approach, which implements random sampling as its query strategy. For the initial training of the model, 300 samples of the EU network (source domain) are used, and then by AL additional 25 samples are added to the training set from the pool of 700 samples of the EU network. After training the base model, it is used as a pre-trained model for the DNN model of the US network (target domain).

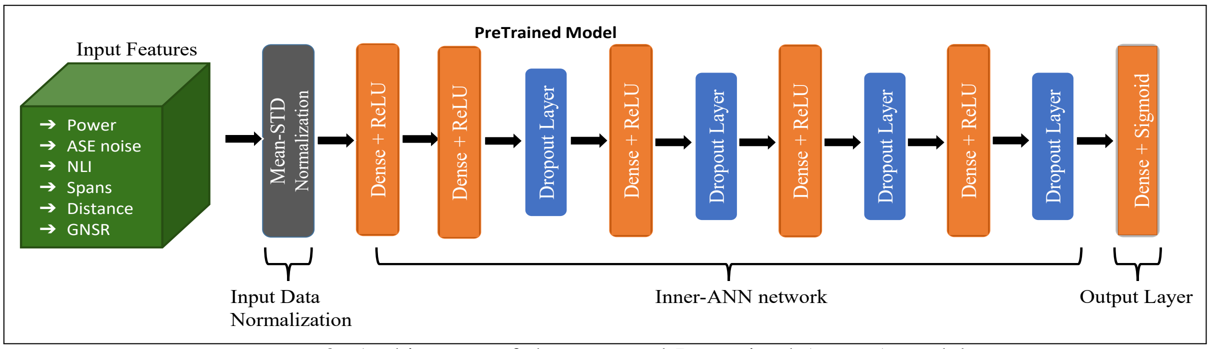

Figure 2: Architecture of the proposed Pre-trained (source) model

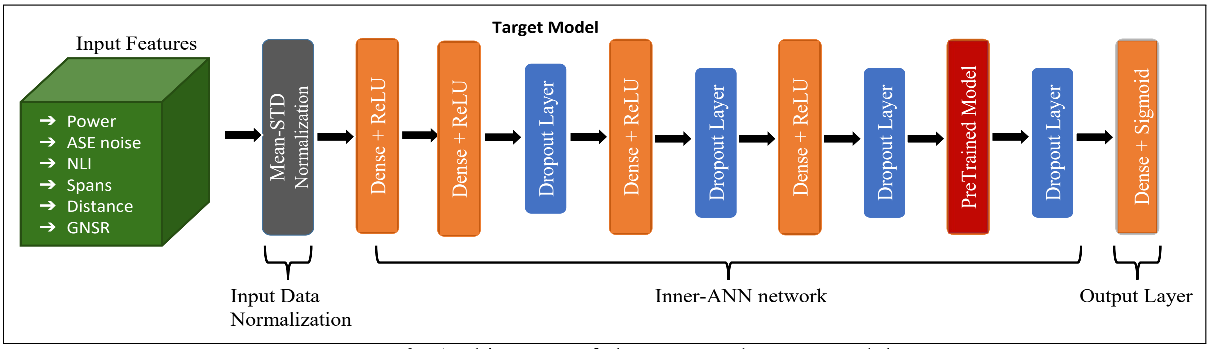

Figure 3: Architecture of the proposed target model

The TL approach is used here as it will enable the pre-trained model to assist the new model by transmitting information gained from a similar previous problem scenario. The Feature extraction (FE) approach of TL is applied here, as it is the most effective domain of TL. The architecture of the target model is illustrated in Fig. 3. As we can see in the proposed targeted architecture (Fig. 3), after combining the TL scheme, four hidden layers are added on top of the pre-trained model along with a new output layer that uses sigmoid as an activation function. A dropout layer after each hidden layer is added to prevent overfitting. This target domain model is then trained with AL to overcome the cost of labeling data samples, and it prevents the issue of fewer data samples available for training with the use of TL in the target domain. For the initial training of the model, initially, 100 data samples are used and 25 additional samples from the pool of 700 samples through AL are added to the training data set. To better understand the proposed ATL's effectiveness, the partial ATL is also applied where the target domain DNN model is trained with TL along with AL, but the pre-trained model (source domain) is trained with a large number of labeled samples without AL. The pre-trained model of partial ATL is trained on 3000 samples, while for the DNN model of the target domain, the same training samples (100 samples) are used for initial training, and then 25 samples are chosen from the pool of 700 samples through AL. A major drawback of the partial ATL approach is that a sufficiently large number of labeled samples are needed for the pre-trained model, which is a costly operation that we prevent in the given scenario. A detailed summary of both ATL and partial ATL is provided in Tab. III.

4. Results and Discussion

We consider the accuracy metric to evaluate the performance of the proposed ATL approach for the LP QoT estimation use case. In Fig. 4, the accuracy score is plotted against the number of samples; during the AL process, the batch size of the samples is set to 10, and it is observed that with an increase in the number of informative data samples in the training dataset, the accuracy of the target model increases gradually. Fig. 5 plots the accuracy score against the number of epochs for DNN training. During this simulation, the batch size of AL was set to 1 and the number of iterations to 25, so in total, the training dataset has 125 data samples

TABLE III: Summary of ATL Vs. Partial ATL.

		ATL	**Partial ATL**
Source domain	Pretrained model	present	present
	AL-based Pretrained model	present	absent
	Pretrained model training sample	300 + 25 additional	3000
Target domain	TF model	present	present
	Pretrained model + TF model	present	present
	TF model training sample	100 + 25 additional	100 + 25 additional

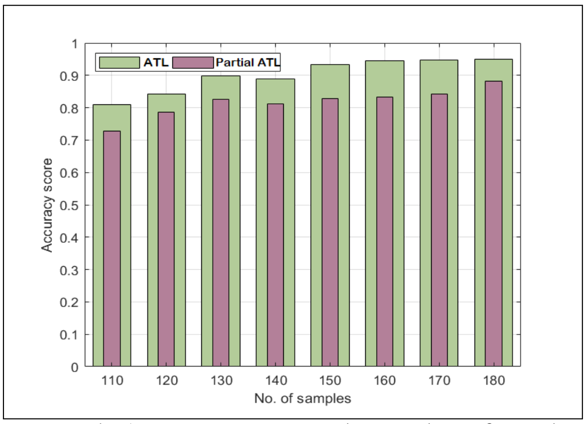

Figure 4: Accuracy score vs. the number of epochs

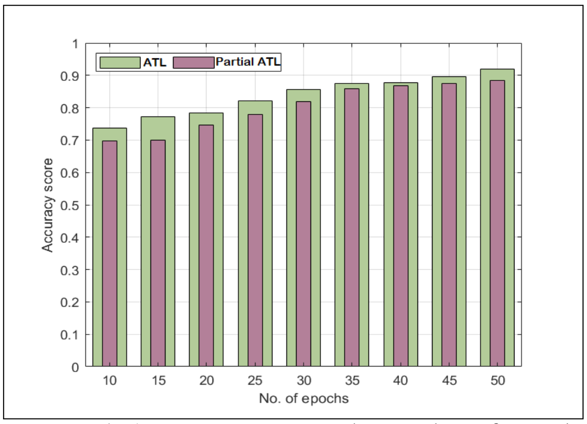

Figure 5: Accuracy score vs. the number of samples

after the completion of AL cycles. The number of epochs varied from 10 to 50 with a difference of 5 epochs, and emphasizes that there is a direct relation between the number of epochs and accuracy score, as with the increase in epochs, the accuracy increases. The model trained by ATL achieves an accuracy of 0.95 with total data samples of 180, and the model trained by partial ATL achieves an accuracy of 0.88 with 180 data samples. During this analysis, the number of epochs was set to 15 with a default batch size of 32. Our proposed pre-trained model has an accuracy score of 0.9078 yet is trained on 325 samples in total, while in the second approach of partial ATL, the pre-trained model is trained on 3000 samples and has an accuracy score of 0.8223. These models were tested on 6000 data samples. This emphasizes that AL is an effective technique in our case. The results shown in Fig. 4 and Fig. 5 demonstrate that ATL performs more efficiently than partial ATL. The primary inadequacy of partial ATL is that for its pre-trained model training, a sufficiently large number of labeled samples are needed, which is a costly operation, and in some cases, it is impossible to achieve such a massive dataset as the one we considered in this work.

5. Conclusion

For a precise LP QoT assessment, optical networks have increasingly adopted machine learning; however, the two major issues in this area are the requirement of a substantial size of training data and efficient domain adaptability after training. This work proposes a hybrid ATL approach comprising transfer learning and active learning strategies. The ATL approach enables high accuracy while using a limited number of training samples, which prevents the cost of labeled data in parallel; it is more adaptive as ATL utilizes the transfer learning to transfer knowledge from one domain to another. Our simulation results validate the effectiveness of the proposed approach for the given LP QoT classification scenario.

REFERENCES

[1] F. Usmani, I. Khan, M. U. Masood, A. Ahmad, M. Shahzad, and V. Curri, "Convolutional neural network for quality of transmission prediction of unestablished lightpaths," MOTL **63**, 2461–2469 (2021).
[2] I. Khan, M. Bilal, and V. Curri, "Assessment of cross-train machine learning techniques for qot-estimation in agnostic optical networks," OSA Continuum **3**, 2690–2706 (2020).
[3] T. Panayiotou, S. P. Chatzis, and G. Ellinas, "Performance analysis of a data-driven quality-of-transmission decision approach on a dynamic multicast-capable metro optical network," JOCN **9**, 98–108 (2017).
[4] F. Usmani, I. Khan, M. U. Masood, A. Ahmad, M. Shahzad, and V. Curri, "Transfer learning aided qot computation in network operating with the 400zr standard," in *ONDM,* (2022), pp. 1–6.
[5] H. Hino, "Active learning: Problem settings and recent developments," (2020).
[6] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang, "A survey of deep active learning," (2020).
[7] A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. L. Rouzic, J. Kundrát, G. Grammel, G. Galimberti, and V. Curri, "Gnpy: an open source application for physical layer aware open optical networks," JOCN **12**, C31–C40 (2020).
[8] I. Khan, M. Bilal, M. Umar Masood, A. D'Amico, and V. Curri, "Lightpath qot computation in optical networks assisted by transfer learning," JOCN **13**, B72–B82 (2021).
[9] V. Curri, "Software-defined wdm optical transport in disaggregated open optical networks," in *ICTON,* (IEEE, 2020), pp. 1–4.
[10] A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. Le Rouzic, J. Kundrat, G. Grammel, G. Galimberti, and V. Curri, "Gnpy: an open source application for physical layer aware open optical networks," JOCN **12** (2020).