Overview of ML-aided QoT Estimation in Optical Networks: A Perspective of Model Generalization

Zhenlin Ouyang[1], Xiaokang Chen[1], Zhengyong Liu[1], Xiaoliang Chen[2], Zuqing Zhu[2]

[1]School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China

[2]School of Information Science and Technology, University of Science and Technology of China, Hefei, China

***Abstract*—The past decade has witnessed a tremendous stride toward automated and intelligent optical networking thanks to the revolutionary development in machine learning (ML). Among the various ML applications for optical networks, quality-of-transmission (QoT) estimation outstands as a fundamental yet challenging task, and therefore, has grabbed intensive research interests. This paper provides an overview of ML-aided QoT estimation. We first describe several representative QoT estimation models. Then, we elicit challenges related to model generalization ability and review the state of the art in this perspective.**

***Index Terms*—Quality-of-transmission (QoT) estimation, machine learning, model generalization.**

I. Introduction

Quality-of-transmission (QoT) estimation is an imperative task for securing the correct operations of optical networks. Accurate QoT estimation facilitates early detection of performance degradations by constant monitoring of lightpaths' health degrees while allowing for allocation of just enough margins to improve resource efficiency. As optical networks scale both horizontally (e.g., larger topologies of multiple heterogeneous domains) and vertically (adoption of novel technologies such as elastic optical networking and space-division multiplexing), this task becomes increasingly challenging and remains a focal point for the research community.

Traditional QoT estimation methods generally fall into two categories: the split-step Fourier method (SSFM) and the Gaussian noise (GN) model. The SSFM method is an analytical model that calculates lightpath QoT through complex iterative computations. This method offers the highest accuracy but at the cost of high computational complexity, making it impractical for real-time applications. The GN model simplifies the mathematical calculations by using statistical methods. While this approach reduces computation time, it trades off the prediction accuracy and necessitates increased provisioning margins to account for the prediction errors.

Recently, thanks to the significant advancements in programmable data-plane technologies, the application of AI in optical networks has become more convenient and efficient [1].The reviving of artificial intelligence (AI) and machine learning (ML) has brought up new opportunities for meeting the aforementioned challenges by exploring data-driven approaches. AI/ML enables learning complex functions directly from data without explicit programming or understanding of the inherent principles of the target systems. This makes AI/ML particularly useful for QoT estimation tasks, as revealed by numerous previous studies.

However, the data-driven nature of ML-aided QoT estimation designs makes their performance highly dependent on the amount and quality of data. Consequently, ML-based designs often suffer from scalability or generalization issues, in circumstances where collecting a large amount of optical performance monitoring (OPM) data is expensive or network conditions are intricate and dynamic. Enhancing model generalization ability thus plays a vital role in paving the path toward the practical deployment of ML-aided QoT estimators in optical networks. Potential solutions include transfer learning approaches that ease the training effort on new tasks by reusing knowledge learned from existing tasks, active learning approaches that optimize the efficiency of data collection, and meta learning designs aiming at extracting generalizable knowledge across tasks. In this paper, we overview ML-aided QoT estimation, and particularly, review the state of the art from the perspective of model generalization.

II. ML-aided QoT Estimation

QoT estimation underpins the following network control and management tasks: *i*) network design and planning, where QoT estimation can help determine the optimal routes, wavelength allocation, amplifier locations, and so on; *ii*) performance monitoring and assurance, where QoT estimation enables real-time and cost-effective monitoring of signal transmission quality and proactive adjustment of lightpath configurations to secure consistent performance levels; *iii*) fault diagnosis and recovery, where QoT estimation can help identify fault locations and develop effective recovery strategies. The most common application of QoT estimation lies in network design and planning, in particular, to estimate the QoT and thereby the feasibility of the unestablished lightpaths. QoT estimation can be realized in the form of classification or regression. Classification involves classifying whether relevant communication quality indicators [such as optical signal-to-noise ratio (OSNR), and bit error rate (BER)] meet the transmission requirements according to preset thresholds. Regression, on the other hand, directly predicts the values of communication quality indicators. Metrics like mean squared error (MSE) are

This work was supported in part by the National Key Research and Development Program of China (2020YFB1806401), National Natural Science Foundation of China (NSFC) (62201627), Guangzhou Basic and Applied Basic Research Foundation (SL2022A04J01907) and the Guangdong Program (2021QN02X039).

Corresponding author: Xiaoliang Chen (email: xlichen@ieee.org).

generally used to calculate the discrepancy between predicted and actual values, and in turn to measure model performance.

When performing QoT estimation, a model requires input features that can characterize a particular lightpath. To ensure that the data can adequately describe the lightpath and that the trained model has strong representation capabilities, researchers in previous studies have opted to use multiple end-to-end communication features. Some of the features chosen by researchers are shown in Table I.

TABLE I
QoT FEATURES

Features	Description
Num of links [km]	Total links that a lightpath traverses, a link consists of several spans
Num of spans [km]	Total spans of a lightpath
Total length [km]	The length of a lightpath, representing the communication distance between two end nodes
Average link length [km]	Average distance of links used by a lightpath
Maximum link length [km]	Distance of the longest link
Launch Power [dBm]	Signal power injected into the network, affecting the communication distance and NLI
Modulation format	Determines a signal's tolerance to noise
Bit rate [bit/s]	Affects a signal's tolerance to noise

Many studies have been devoted to applying ML to QoT estimation in optical networks, achieving promising results, as summarized in Table II. K-nearest neighborhood (KNN) is one of the simplest supervised learning methods. It classifies samples by calculating the distances between samples (using the input features) and by assigning a sample the most frequent label within the K nearest neighbors. In [2], the authors used $K = 1$ and $K = 10$, and examined the prediction results under the Euclidean distance, cosine distance, and weighted Euclidean distance. The final results showed that the highest prediction accuracy was achieved with $K = 1$, and among the different distance metrics, the weighted Euclidean distance performed the best. The authors also used a support vector machine (SVM) for QoT estimation. SVM finds the optimal hyperplane through kernel functions for regression, which can be seen as a generalization of KNN. The authors considered linear, quadratic, and cubic SVM, along with Gaussian kernels with $\gamma = 1$ and $\gamma = 3$. The final results indicated that cubic SVM performed the best among all cases. In [3], the authors used the random forest method for QoT estimation. The random forest algorithm consists of multiple decision trees, each analyzing the relationship between input features and results to form the corresponding model, thus classifying the input samples. In the random forest algorithm, each decision tree randomly selects features to ensure diversity and produces a result that is then determined by majority voting. The authors considered numbers of estimators of 1, 5, 25, 100, and 500. The final results showed that more estimators resulted in better performance but also increased training and prediction time.

Neural networks (NNs) are the most popular ML algorithms in recent years. An NN fits complex problems through training data iteratively and can work in an end-to-end manner. In [4], the authors used NNs for QoT estimation and proposed a time-sequence QoT estimation model based on recurrent neural networks (RNNs). The results showed that NNs achieved the highest prediction accuracy of 99.56% when using all features, and the accuracy decreased as features were reduced. When considering time sequences, long short-term memory (LSTM), gated recurrent unit (GRU), and encoder-decoder LSTM models were used. The results indicated that GRU performed the best, and more data are needed to train higher-performance models.

In [5], the authors proposed a convolutional neural network (CNN) based network-wide QoT estimation method to improve the prediction efficiency in large-scale deployments. They utilized CNN's ability to extract network-level feature information from network matrices for prediction. The results showed that QoT estimation accuracy exceeding 98% across different network topologies was achieved. Compared to CNN, graph neural networks (GNNs) can extract features from irregular graphs such as network topology, thus gaining significant attention in recent years. In [6], the authors used deep graph convolutional neural network (DGCNN) to model QoT problems through subgraphs and trained GNNs using the features of in-service lightpaths' graph data. This method not only predicted QoT for unestablished lightpaths but also inferred whether new lightpath deployments would affect existing ones. The results showed that GNNs achieved over 92% accuracy, while that of traditional DNNs only reached 77%.

However, we can find that all the aforementioned ML methods require the same initial step, which is the use of data to train the model. For simpler models like KNN, SVM, and RF, less training data are required, but their ability to fit problems is also weaker. Meanwhile, for models like NNs, the more complex the model is, the better representation capability they have. However, a more complex model also implies more data for training to ensure performance. Specifically, when the data distribution of the actual application domain differs from that of the source domain, a model's expected performance can significantly decrease. In practice, researchers aim to train sufficiently powerful models to enhance problem representation capabilities (for QoT, this means improving prediction accuracy) while ensuring the model's generalization ability to different data distributions (i.e., unseen conditions). This remains a significant challenge for ML-aided QoT estimation that hinders its practical application.

III. MODEL GENERALIZATION

In practical optical networks, data collection on-site requires the deployment of multiple OPM devices or modules and telemetry services that constantly stream OPM data to the centralized control plane, which incurs significant costs. On

TABLE II
METHODS FOR QOT ESTIMATION

Methods	Classification/Regression	Metrics	Feature	Result	Ref.
KNN	Classification	OSNR	Num of hops; num of spans; total length; average link length; maximum link length; average span attenuation; average dispersion; modulation format	Accuracy=96.47% when $K = 1$ and accuracy=92.93% when $K = 10$ using Euclidean distance	[2]
Random Forest	Classification	BER	Num of links; total length; longest link length; traffic volume; modulation format	Using $1, 5, 25, 50, 100, 500$ estimators, and the accuracy is 92% with one estimator and 96% with five types.	[3]
SVM	Classification	OSNR	Num of hops; num of spans; total length; average link length; maximum link length; average span attenuation; average dispersion; modulation format	3D SVM has better performance (accuracy=96.62%) than others while Gaussian kernel $\gamma = 3$ is better than $\gamma = 1$ (93.05% and 91.79%, respectively)	[2]
ANN, RNN	Classification	SNR	link length; span length; number of spans; modulation format; channel power; data rate	The accuracy of ANN is 99.56% with 0.276 ms computation time and the RNN's prediction performance degrades with the forecast horizon increase	[4]
DCNN	Classification	SNR	Length; Num of spans; BER; modulation format	Using two matrix to represent the feature and estimate the network-wide QoT, best accuracy is 99.52%	[5]
GNN	Classification	SNR	Total length; longest length; starting slot; num of slots; modulation format; num of EDFA; num of links; num of adjacent links; overall slots; established connections' BER	Using the information of the in-service lightpath around the unestablished lightpath. The accuracy of the GNN is between 92% and 100%, which outperforms the DNN	[6]

the other hand, using synthetic data raises concerns about the deficiency of the dataset‘s diversity and the generalization ability of the models trained on such data. Under these circumstances, developing a model that performs well while having strong generalization capabilities for unseen conditions is an urgent challenge for applying ML in optical networks and many efforts have been devoted to improving model generalization. In this section, we review several works dedicated to improving the generalization ability of ML-based QoT estimators.

A. Transfer Learning

Transfer learning is an ML technique where a model developed for a particular task (source task) is reused as the starting point for a model on a different task (target task). This method is particularly useful when the target tasks have limited data, as it allows to leverage the knowledge gained from the source tasks to improve performance on the target tasks.

Transfer learning is particularly well-suited to the characteristics of QoT estimation tasks in optical networks. Consequently, numerous studies have focused on utilizing transfer learning to improve model generalization ability in QoT estimation tasks, such as [7]–[10], [13].

In [7], the authors considered that fluctuations near the operating point of network elements could introduce uncertainty in the nominal values of collected data, leading to generalized signal-to-noise ratio (GSNR) uncertainty and increased system margins. The authors pre-trained a model with in-service network data and then applied transfer learning to reuse the pre-trained model in an unused network environment, thereby reducing GSNR calculation uncertainty. Experimental results showed that the model's output GSNR values fit well with the actual conditions and significantly reduced the discrepancy between predicted GSNR and nominal values.

In [8], the authors proposed an evolutionary algorithm-enhanced transfer learning method for QoT estimation in multi-domain elastic optical networks. They utilized a genetic algorithm to optimize the neural network architecture and set transfer weights based on source and target tasks to improve method performance. After pre-training the model, they attempted to freeze certain structures of the pre-trained model and added untrained hidden layers in the new model to enhance the model's performance. The genetic algorithm was used to find the optimal model architecture for knowledge transfer, maximizing the model's performance increment. The final results showed that, while achieving 95% accuracy, the proposed method could reduce up to $13\times$ training data compared with the traditional methods.

The work in [9] focused on selecting samples from the source and target domain when using transfer learning, thereby improving model performance with less data. The authors designed a sample distribution matching model to filter data from the source domain and find samples that match the target domain's data distribution, enhancing the pre-trained model's performance in the target domain. Different sample matching algorithms were designed and their pros and cons were discussed. The results showed that the proposed method saved fine-tuning samples of up to 28.5% compared with

TABLE III
GENERALIZATION

Methods	Model	Feature	Result	Ref.
Transfer learning	DNN	Power; ASE; NLI; frequency; spans	Trained the model with in-service network data and reduced the GSNR uncertainty of the unused network with transfer learning. The experimental results showed that the average margin was reduced by 0.27 dB.	[7]
Transfer Learning	ANN	Num of links; modulation format; Amplifiers gain	Genetic algorithm was used to search the optimal model architectures and transfer weight, the proposed approach can reduce the training data by up to 13 times compared to the conventional method in the same performance.	[8]
Transfer learning	ANN	Total length; maximum link length; num of hops; num of spans; num of adjacent channels	Proposed a sample-distribution-matching algorithm to find the suitable samples for the target domain, results showed that proposed approach outperforms the traditional transfer learning model and saves more than 28.5% fine-tuning samples	[9]
Transfer learning	ANN	Launch power; wavelength; transmission distance	Proposed a neuron-level transfer learning approach which uses PSO algorithm to determine the trainable and frozen neurons of ANN, the proposed approach achieved better performance than conventional transfer learning.	[10]
Domain adaptation	SVM; logistic regression; RF	Total length; num of traversed links; maximum link length; amount of transmitted traffic; modulation format; traffic volume; guardband size	Proposed a domain adaptation approach to compare with the approach that simply joins the target and source domain data together, the domain adaptation approach outperformed the normal ML method and decreased the dataset distribution.	[11]
Active learning	Gaussian processes	Num of links; total length; maximum link length; traffic volume; modulation format	Active learning were used to search for the next training samples. The training data were decreased (at least 5% and up to 75%) to achieve the same performance.	[12]
Active & transfer learning	DNN	Power; ASE noise; NLI; spans; distance; GSNR	Active learning were used when choosing the data on both the pre-training and fine-tuning stages. Results showed that active learning enhanced the performance of the model when using transfer learning	[13]
Domain adaptation & active learning	Gaussian processes	Total length; maximum link length; num of traversed links; amount of transmitted traffic; traffic volume; modulation format; guardband size	Domain adaptation and active learning were used for QoT estimation and made a comparison between these two approaches. The active learning method requires a few dozen samples, whereas a few hundred samples are needed for domain adaptation	[14]
Meta learning	ANN	Num of spans; Maximum span length; average span length; signal power; link length; chromatic dispersion; NLI parameter, etc.	Meta learning-based approach has better performance and robustness in the presence of different parameter uncertainties. With parameter uncertainty increasing, the RMSE of the traditional model was 1.22 dB while that of the meta learning model was 0.83 dB.	[15]
Meta Learning	ANN	Total length; longest link length; num of spans; num of hops; wavelength; adjacent channels	Conventional training, transfer learning, and meta learning were compared in parallel. The results showed that meta learning has better data-saving ability compared with transfer learning (by 55%).	[16]
Meta Learning	CNN	Constellation grayscale maps	Proposed an auxiliary task for meta learning to enhance the training convergence. Modulation format identification accuracy reached 100% as an auxiliary task and QoT estimation's MSE was 0.18 dB.	[17]
Continual Learning	Invariant CNN	Channel power; noise figure of spans; EDFA gain of spans; Span length	Enabled variation of link feature parameter for QoT estimation using invariant CNN. A joint training algorithm was proposed to alleviate the time and feature-length dependency while continual learning framework was used to perform the highest efficiency and lowest training cost of QoT estimation	[18]
Composable ML	ANN, RNN	Launch power; symbol rate; wavelength; channel loading; physical length; attenuation coefficient; amplifier gain; noise figure	Proposed a composable ML method to generalize different tasks in the same domain. Results showed that MAE below 1.03 dB was achieved	[19]

traditional transfer learning methods.

Another evolutionary transfer learning method was proposed by [10]. Compared to [8], the authors of [10] proposed neuron-level transfer learning. This method used a particle swarm optimization (PSO) algorithm to determine the trainable and frozen neurons of an NN, rather than layers of NN, as in traditional transfer learning methods. The results showed that the proposed method achieved higher accuracy compared with traditional layer-level transfer learning, and also improved the model's reliability and throughput in the application phase.

Domain adaptation (DA) is a branch of transfer learning that focuses on improving the performance of a model on a target domain by leveraging knowledge from a related source domain. This technique is particularly useful when there is a shift in the distribution of data between the source and target domain. In [11], the authors proposed to use DA techniques on target datasets to address the issue of data scarcity or unavailability. They introduced two DA methods based on different available data and compared them with ordinary models. The results showed that these two methods performed better than simply joining the target domain and source domain data.

Domain adaptation and active learning are both methods to enhance performance on small datasets. In [14], the authors compared the two methods and showed that the two methods outperformed traditional ML methods on limited datasets. As the number of samples increased, the prediction performance of both methods improved significantly, with active learning requiring fewer samples than domain adaptation.

A composable ML framework has been proposed in [19], aiming at generalizing ML-aided cognitive applications. The authors introduced three fundamental functional modules corresponding to initialization, feature extraction, and model inference, trained in an end-to-end manner. The authors found that composable ML can be generalized across different tasks within the same domain, such as QoT estimation for different lightpaths. Experimental results demonstrated the robust generalization ability of composable ML, achieving a mean absolute error (MAE) of only 1.06 dB in QoT estimation tasks.

Transfer learning and domain adaptation both aim to leverage the common knowledge from different data distributions to train models. By learning the shared knowledge between different distributions, they reduce the need for extensive training time and data in other specific distributions, thus improving the generalization of machine learning (ML) models. Typically, researchers combine various strategies with transfer learning or domain adaptation, such as selective sampling of data, choosing optimal model architectures, or refining training methods. This combination of techniques enables the models to perform better across diverse environments while requiring fewer labeled examples.

B. Active Learning

Active learning is a type of ML where the algorithm selects the most informative data points from which to learn. This approach is particularly useful when labeled data is scarce or expensive to obtain, as it aims to maximize model performance with a minimal amount of labeled data.

Due to the need to deploy probes for data collection in optical networks, the cost of data collection is expensive. In [12], to reduce the number of probes in optical networks, the authors used active learning to train Gaussian processes for QoT estimation. Using this method, the QoT estimator can iteratively select the most suitable samples for model training, thereby training a high-performance model with the fewest samples. By using active learning, the probability distribution output by the Gaussian processes is input into an acquisition function, which is then used to find the most appropriate next training sample. The results showed that this method could significantly reduce the number of training samples (by at least 5% and up to 75%) while achieving an accuracy comparable to that of conventional offline ML methods.

Active learning and transfer learning were combined in [13], which not only reduces the number of samples needed for model training but also enables the model to be used in networks with different configurations. The authors pointed out that directly training with limited labeled data could lead to overfitting and degrade model performance. Their proposed method uses active learning to find the optimal high-quality samples for model training in both the pre-training and fine-tuning stages and then applies the pre-trained model to QoT estimation tasks in other networks. They compared the effectiveness of transfer learning without using active learning to select samples during the fine-tuning stage. The final results showed that using active learning throughout both stages was more effective than using it only in the pre-training stage, demonstrating the effectiveness of active learning in both stages.

Different approaches to enhance model generalization and data-use efficiency have been explored. In [18], the authors identified two challenges faced by ML-based QoT estimation: the varying number of transmission feature parameters and data distribution drift. To address these challenges, they proposed an invariant CNN-based estimation model. This model encodes link parameters of different lengths into a fixed-length representation and quickly adapts to different data distributions using limited data. The authors introduced a joint training algorithm to mitigate the model's dependence on time and features and employed a continual learning workflow to achieve maximum efficiency in QoT estimation with minimal training cost. The results demonstrated that the proposed method's 99th percentile prediction error average was less than 1.0 dB and showed stronger stability compared with benchmarks over different periods.

Active learning is machine learning focused on selecting the most efficient training samples. Through sample selection algorithms, it identifies and labels the most suitable samples for the next training iteration, maximizing training efficiency. Compared to passively using large amounts of randomly selected data, active learning significantly reduces the number of training samples required. Many researchers employ active learning frameworks to adaptively choose samples for real-time training in different scenarios, making it an effective strategy to enhance model generalization.

C. Meta Learning

Meta learning, also known as "learning to learn", aims to develop models that can learn new tasks more efficiently by leveraging prior experience. Instead of focusing on learning to perform a single task, meta-learning algorithms aim to acquire knowledge that can be generalized across multiple tasks. Meta learning is widely used for few-shot learning enabling models to achieve considerable performance with very few samples available. This characteristic is highly applicable in the data-scarce environment of optical networks. Consequently, researchers are dedicated to using meta learning to improve model generalization under conditions of data scarcity.

In [15], the authors aimed to address the issue of parameter uncertainty in physical layer impairment modeling by using online adaptation. They sought to improve the robustness of the model and its effective adaptation to limited real system data through meta learning. The authors experimented with two meta learning algorithms, model-agnostic meta-learning (MAML) and Reptile, and compared their performance with traditionally trained models. The results showed that traditional models performed the worst under all levels of parameter uncertainty, while meta learning methods maintained stability as uncertainty increased.

Since both transfer learning and meta learning address the issue of limited data, the authors in [16] compared ordinary NN, transfer learning based, and meta learning based methods. The results demonstrated that, when pre-trained on one network and applied to different networks with scarce data, both transfer learning and meta learning outperformed ordinary methods. Furthermore, when comparing transfer learning with meta learning, the experiments revealed that to achieve the same model performance, meta learning could reduce the data size by approximately 55% compared with that required by transfer learning. This indicates that transfer learning requires more data for sufficient knowledge transfer, whereas meta learning exhibits a stronger data generalization ability during the pre-training stage.

In [17], the authors introduced the concept of auxiliary tasks to help meta learning converge faster and compared standard meta learning with meta learning incorporating auxiliary tasks. In their experiments, they used modulation format identification as the auxiliary task to improve the convergence of the QoT model training. The results showed that with the auxiliary task, the accuracy of modulation format identification could reach 100%, and the lowest QoT estimation MSE could be 0.18 dB. Additionally, in the fine-tuning stage, only four samples were needed to achieve remarkable generalization ability and adaptability.

Meta learning can be viewed as a specialized training approach that enables a model to achieve rapid convergence using a small number of training samples in specific tasks, allowing it to learn distinctions between different tasks. Researchers leverage its fast convergence and multi-task compatibility to enhance model generalization. This makes meta learning particularly effective in improving the performance of models when faced with varied tasks and limited data, which is critical in the dynamic and data-scarce environment of optical networks.

Conclusion

Improving model generalization ability is a crucial pathway for the practical deployment of ML applications in optical networks. In this article, we summarized recent efforts on ML-aided QoT estimation that particularly focus on model generalization, leveraging techniques such as transfer learning, active learning, and meta learning.

The generalization of models has consistently been a critical challenge for the practical deployment of machine learning (ML) methods in optical networks. Ensuring model generalization is essential for transitioning from experimental phases to in-filed applications. First, in current research, most simulation and field experiments are conducted on similar or identical network environments, which presents significant challenges for enhancing the model's generalizability. In future studies, researchers could explore novel methods to ensure that ML models can meet experimental expectations across various network environments and data distributions. Secondly, in the context of automated and intelligent optical networks, reducing computational resources and time costs is an important consideration. One potential direction is to investigate the use of multimodal approaches or other models that can simultaneously address multiple tasks (e.g., fault management, QoT prediction, routing, spectrum assignment, and others), exploring ways to develop a single model capable of solving multiple problems efficiently.

References

[1] B. Niu, J. Kong, S. Tang, Y. Li, and Z. Zhu, "Visualize Your IP-Over-Optical Network in Realtime: A P4-Based Flexible Multilayer In-Band Network Telemetry (ML-INT) System," *IEEE Access*, vol. 7, pp. 82 413–82 423, 2019.

[2] R. M. Morais and J. Pedro, "Machine Learning Models for Estimating Quality of Transmission in DWDM Networks," *J. Opt. Commun. Netw.*, vol. 10, no. 10, p. D84, Oct. 2018.

[3] C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-Learning Method for Quality of Transmission Prediction of Unestablished Lightpaths," *J. Opt. Commun. Netw.*, vol. 10, no. 2, p. A286, Feb. 2018.

[4] S. Aladin, A. V. S. Tran, S. Allogba, and C. Tremblay, "Quality of Transmission Estimation and Short-Term Performance Forecast of Lightpaths," *J. Lightw. Technol.*, vol. 38, no. 10, pp. 2807–2814, May 2020.

[5] P. Safari, B. Shariati, G. Bergk, and J. K. Fischer, "Deep Convolutional Neural Network for Network-wide QoT Estimation," in *Proc. Conf. Opt. Fiber Commun. (OFC)*, 2021, p. Th4J.3.

[6] G. Savva, T. Panayiotou, I. Tomkos, and G. Ellinas, "Deep Graph Learning for QoT Estimation of Unseen Optical Sub-Network States: Capturing the Crosstalk Impact on the In-Service Lightpaths," *J. Lightw. Technol.*, vol. 40, no. 4, pp. 921–934, Feb. 2022.

[7] I. Khan, M. Bilal, M. Umar Masood, A. D'Amico, and V. Curri, "Lightpath QoT Computation in Optical Networks Assisted by Transfer Learning," *J. Opt. Commun. Netw.*, vol. 13, no. 4, p. B72, Apr. 2021.

[8] C.-Y. Liu, X. Chen, R. Proietti, and S. J. B. Yoo, "Performance Studies of Evolutionary Transfer Learning for End-to-End QoT Estimation in Multi-domain Optical Networks [Invited]," *J. Opt. Commun. Netw.*, vol. 13, no. 4, p. B1, Apr. 2021.

[9] Z. Gu, T. Qin, Y. Zhou, J. Zhang, and Y. Ji, "Sample-distribution-matching-based Transfer Learning for QoT Estimation in Optical Networks," *J. Opt. Commun. Netw.*, vol. 15, no. 9, p. 649, Sep. 2023.

[10] Y. Zhou, Z. Gu, J. Zhang, and Y. Ji, "Evolutionary Neuron-level Transfer Learning for QoT Estimation in Optical Networks," *J. Opt. Commun. Netw.*, vol. 16, no. 4, p. 432, Apr. 2024.

[11] C. Rottondi, R. Di Marino, M. Nava, A. Giusti, and A. Bianco, "On the Benefits of Domain Adaptation Techniques for Quality of Transmission Estimation in Optical Networks," *J. Opt. Commun. Netw.*, vol. 13, no. 1, p. A34, Jan. 2021.

[12] D. Azzimonti, C. Rottondi, and M. Tornatore, "Reducing Probes for Quality of Transmission Estimation in Optical Networks with Active Learning," *J. Opt. Commun. Netw.*, vol. 12, no. 1, p. A38, Jan. 2020.

[13] H. Tariq, F. Usmani, I. Khan, M. U. Masood, A. Ahmad, and V. Curri, "Iterative Transfer Learning Approach for QoT Prediction of Lightpath in Optical Networks," in *Int. Conf. Transparent Opt. Netw. (ICTON)*, Jul. 2023, pp. 1–4.

[14] D. Azzimonti, C. Rottondi, A. Giusti, M. Tornatore, and A. Bianco, "Comparison of Domain Adaptation and Active Learning Techniques for Quality of Transmission Estimation with Small-sized Training Datasets [Invited]," *J. Opt. Commun. Netw.*, vol. 13, no. 1, p. A56, Jan. 2021.

[15] X. Liu, H. Lun, L. Liu, Y. Zhang, Y. Liu, L. Yi, W. Hu, and Q. Zhuge, "A Meta-Learning-Assisted Training Framework for Physical Layer Modeling in Optical Networks," *J. Lightw. Technol.*, vol. 40, no. 9, pp. 2684–2695, May 2022.

[16] S. Lin, Z. Gu, J. Zhang, and Y. Ji, "Meta Learning Based QoT Estimation of lightpaths with few samples for Optical Networks," in *Proc. Aisa Commun. Photon. Conf. (ACP)*, Nov. 2023, pp. 1–4.

[17] Y. Zhang, P. Zhou, Y. Liu, J. Wang, C. Li, and Y. Lu, "Fast Adaptation of Multi-task Meta-learning for Optical Performance Monitoring," *Opt. Express*, vol. 31, no. 14, p. 23183, Jul. 2023.

[18] Q. Wang, Z. Cai, A. P. T. Lau, Y. Li, and F. Nadeem Khan, "Invariant Convolutional Neural Network for Robust and Generalizable QoT Estimation in Fiber-optic Networks," *J. Opt. Commun. Netw.*, vol. 15, no. 7, p. 431, Jul. 2023.

[19] H. Gao, X. Chen, C. Lu, and Z. Li, "On the Generalization of Cognitive Optical Networking Applications Using Composable Machine Learning," *J. Opt. Commun. Netw.*, vol. 16, no. 6, p. 631, Jun. 2024.