

Article

Self-Attention Mechanism-Based Multi-Channel QoT Estimation in Optical Networks

Yuhang Zhou ¹, Xiaoli Huo ², Zhiqun Gu ^{1,*}, Jiawei Zhang ¹, Yi Ding ², Rentao Gu ¹  and Yuefeng Ji ¹

¹ State Key Lab of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China

² China Telecom Research Institute, Beijing 102209, China

* Correspondence: guzhiqun@bupt.edu.cn

Abstract: It is essential to estimate the quality of transmission (QoT) of lightpaths before their establishment for efficient planning and operation of optical networks. Due to the nonlinear effect of fibers, the deployed lightpaths influence the QoT of each other; thus, multi-channel QoT estimation is necessary, which provides complete QoT information for network optimization. Moreover, the different interfering channels have different effects on the channel under test. However, the existing artificial-neural-network-based multi-channel QoT estimators (ANN-QoT-E) neglect the different effects of the interfering channels in their input layer, which affects their estimation accuracy severely. In this paper, we propose a self-attention mechanism-based multi-channel QoT estimator (SA-QoT-E) to improve the estimation accuracy of the ANN-QoT-E. In the SA-QoT-E, the input features are designed as a sequence of feature vectors of channels that route the same path, and the self-attention mechanism dynamically assigns weights to the feature vectors of interfering channels according to their effects on the channel under test. Moreover, a hyperparameter search method is used to optimize the SA-QoT-E. The simulation results show that, compared with the ANN-QoT-E, our proposed SA-QoT-E achieves higher estimation accuracy, and can be directly applied to the network wavelength expansion scenarios without retraining.

Keywords: quality of transmission (QoT) estimation; nonlinear effect; multi-channel; self-attention mechanism



Citation: Zhou, Y.; Huo, X.; Gu, Z.; Zhang, J.; Ding, Y.; Gu, R.; Ji, Y. Self-Attention Mechanism-Based Multi-Channel QoT Estimation in Optical Networks. *Photonics* **2023**, *10*, 63. <https://doi.org/10.3390/photonics10010063>

Received: 31 October 2022
Revised: 27 November 2022
Accepted: 3 January 2023
Published: 6 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Precise and fast estimation of quality of transmission (QoT) for a lightpath prior to its deployment becomes capital for network optimization [1]. Traditional physical layer model (PLM)-based QoT estimation utilizes optical signal transmission theories to predict the lightpaths' QoT values, which mainly includes two approaches: (1) sophisticated analytical models, such as the split-step Fourier method [2]; (2) approximated analytical models, such as the Gaussian Noise (GN) model [3]. The former approach analyzes various physical layer impairments and provides high estimation accuracy, but it requires high computational resources. Thus, sophisticated analytical models are unable to estimate the lightpaths' QoT values online and are not scalable to large-scale networks and dynamic network scenarios [2], and the latter approach, of which the widely used GN model simplifies the nonlinear interference (NLI) to additive Gaussian noise, provides quick QoT estimation for lightpaths, but it requires an extra design margin to ensure the reliability of the lightpaths' transmission in the worst-case, thus leading to underutilization of network resources [3]. The PLM-based QoT estimation cannot ensure high accuracy and low computational complexity simultaneously.

Machine learning (ML) has powerful data mining and fast prediction abilities, it has been widely used for QoT estimation. Most of the existing studies focus on the QoT estimation for a signal lightpath [4–20], and all of these achieve high accuracy. For example, in [4], three ML-based classifiers, which include random forest (RF), support vector machine

(SVM), and K-nearest neighbor (KNN), are proposed to predict the QoT of lightpaths. The simulation results show that these classifiers achieve high accuracy and the SVM-based classifier achieves the best performance with a classification accuracy of 99.15%. In [9], a neural network (NN)-based QoT regressor is proposed, and the experimental results show that the regressor achieves a 90% optical signal-to-noise (OSNR) prediction of a 0.25 dB root-mean-squared-error (RMSE) on a mesh network. In [15], a meta-learning assisted training framework for an ML-based PLM is proposed, and the framework can improve the model robustness to uncertainties of the parameters and enable the model to converge with few data samples. In [20], it is explored how ML can be used to achieve lightpath QoT estimation and forecast tasks, and the data processing strategies are discussed with the aim to determine the input features of the QoT classifiers and predictors.

However, due to the nonlinear effect of fibers, the newly deployed lightpaths (new-LPs) degrade the QoT of previously deployed lightpaths (pre-LPs). The single-channel (lightpath) QoT estimation only provides the QoT information of the new-LPs, which leads to the decrease in QoT estimation accuracy of the pre-LPs. Thus, it is necessary to estimate the QoT of new-LPs and pre-LPs simultaneously, i.e., multi-channel QoT estimation. In [21], a deep grapy convolutional neural network (DGCNN)-based QoT classifier is proposed, of which the aim is to accurately classify any unseen lightpath state. In [22], a novel deep convolutional neural network (DCNN)-based QoT classifier is proposed for network-wide QoT estimation. The above works achieve high accuracy of multi-channel QoT classification. Nevertheless, the QoT classifier cannot provide detailed QoT values of lightpaths; thus, it cannot be directly applied to network planning, such as modulation format assignment. Reference [23] proposes an ANN-based multi-channel QoT estimator (ANN-QoT-E) over a 563.4 km field-trial testbed, and the mean absolute error (MAE) is about 0.05 dB for the testing data. However, in optical networks, there exist three NLIs, i.e., self-channel interference (SCI), cross-channel interference (XCI), and multi-channel interference (MCI). When the interfering channel with a higher power is closer to the spectrum of the channel under test (CUT), stronger NLI is introduced on the CUT [3]. Thus, the effects of the interfering channels on the CUTs' QoT are different. The ANN-QoT-E proposed in [23] neglects the different effects of the interfering channels in its input layer, which affects its accuracy. Therefore, how to assign different weights to interfering channels for the QoT estimation of CUTs to enhance the accuracy becomes a crucial problem.

In this paper, we extend our previous work in [24], which applies self-attention mechanisms to improve the accuracy of the ANN-QoT-E. The comparison of this work with the previous works is shown in Table 1, and the main contributions of this paper are summarized as follows:

- (1) We propose a self-attention mechanism-based multi-channel QoT estimator (SA-QoT-E) to improve the estimation accuracy of the ANN-QoT-E, where the input features are designed as a sequence of channel feature vectors, and the self-mechanism dynamically assigns weights to the interfering channels for the QoT estimation of the CUTs.
- (2) We use a hyperparameter search method to optimize the SA-QoT-E, which selects optimal hyperparameters for the SA-QoT-E to improve its estimation accuracy.
- (3) We show the performance of the SA-QoT-E via extensive simulations. The simulation results show that the SA-QoT-E achieves a higher estimation accuracy than the ANN-QoT-E proposed in [23], and it is verified that the assignment of attention weights to the interfering channels conforms to the optical transmission theory. Compared with the ANN-QoT-E, the SA-QoT-E is more scalable, and it can be directly applied to network wavelength expansion scenarios without retraining. By analyzing the computational complexity of the SA-QoT-E and the ANN-QoT-E, it is concluded that the SA-QoT-E has higher computational complexity. However, the training phase of QoT estimators is offline and the computational complexity of the SA-QoT-E is acceptable; thus, the SA-QoT-E still has more advantages than the ANN-QoT-E in practical network applications.

Table 1. Comparison of this work with the previous works.

Work	QoT Classification	QoT Regression	Channel Effect Quantification	Hyperparameters Search
Ref [21]	✓			
Ref [22]	✓			
Ref [23]		✓		
Ref [24]		✓	✓	
This work		✓	✓	✓

The remainder of this paper is organized as follows. In Section 2, we describe the principle of the self-attention mechanism-based multi-channel QoT estimation scheme. The dataset generation and simulation setup are shown in Section 3. In Section 4, we discuss the simulation results in terms of convergence, estimation accuracy, scalability, and computational complexity. Finally, we make a conclusion in Section 5.

2. Self-Attention Mechanism-Based Multi-Channel QoT Estimation

The self-attention mechanism improves the accuracy of the task model by calculating the relevance between Query (Q: query state) and Key (K: candidate state), allowing the task model to dynamically pay attention to the input feature vectors and extract more important information [25]. To solve the problem that ANN models ignore the different effects of the interfering channels in their inputs, we apply the self-attention mechanism to assign the dynamic weights to the input channel features, i.e., SA-QoT-E.

The schematic diagram of the SA-QoT-E is shown in Figure 1a. The steps of the SA-QoT-E are shown as follows:

- (1) Network database collection: The network database, which includes the transmission configuration (such as the launch power of each channel, the wavelength allocation, and transmission distance) and the corresponding QoT value of each channel, is collected from the network topology first.
- (2) Channel features input: Then, the channel features from the network database are input to the self-attention mechanism block.
- (3) New channel features generation: The self-attention mechanism assigns weights to the input channel features according to their effects on the CUT. Thus, the new channel features with interfering channel information (ICI) are generated.
- (4) Channels QoT estimation: Finally, the channel features with ICI are input into an ANN model to estimate the QoT values of the channels.

In the SA-QoT-E, the database generation in step (1) is shown in Section 3 and the ANN model in step (4) is a classic ML model [26]. Thus, in this section, we introduce the design of the input of the self-attention mechanism in step (2) and the process of the self-attention mechanism in step (3).

2.1. The Design of the Input of Self-Attention Mechanism

The self-attention mechanism takes a sequence of feature vectors as input. For a certain feature vector in the sequence, the self-attention mechanism dynamically assigns a weight to each feature vector according to its effect on the vector under test, and finally generates a new feature vector containing the entire sequence information. In optical transmission systems, due to the nonlinear effect of fibers, the characteristics of the interfering channels (such as the launch power and the wavelength) decide their effects on the QoT of the CUT, where the interfering channels and the CUT pass through the same route. Therefore, in the SA-QoT-E, we design the input of the self-attention mechanism as a sequence of channel feature vectors, where the channels pass through the same route and the channel feature vectors contain the corresponding channel transmission characteristics. Each channel feature vector in a sequence is shown as follows:

$$F_i = [p_i, w_i, l_i], \tag{1}$$

where F_i represents the channel feature vector of the i -th channel, p_i is the launch power of the i -th channel, w_i is the wavelength of the i -th channel, and l_i is the transmission distance of the i -th channel.

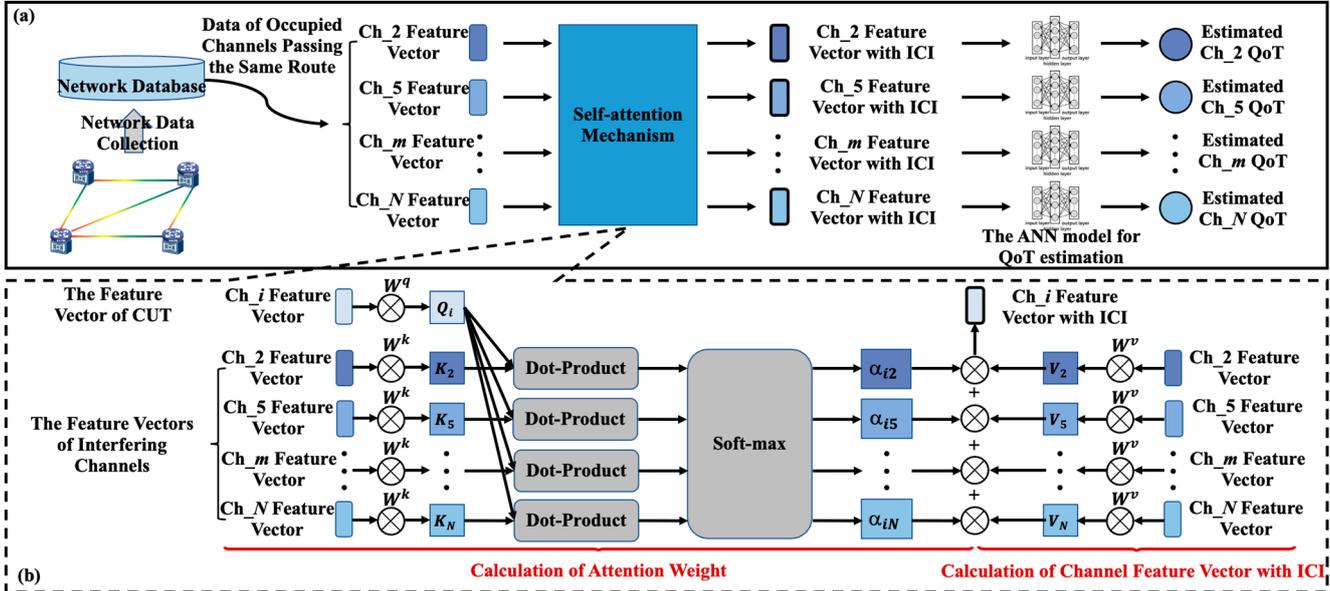


Figure 1. (a) Schematic diagram of self-attention mechanism-based multi-channel QoT estimation scheme; (b) process of self-attention mechanism.

2.2. The Process of Self-Attention Mechanism Block

In the self-attention mechanism of the SA-QoT-E, the attention weights represent the effects of the interfering channels. The higher the attention weight of the interfering channel, the more the SA-QoT-E pays attention to the corresponding interfering channel. There are three parameter matrices trained from the training data, which are W^q , W^k , and W^v , and the channel feature vectors are multiplied by these matrices to obtain the corresponding Q, K, and V. Then, the attention weight is a function of Q and K, and the new channel feature vector with ICI is calculated by the weighted summation of the attention weight and V. The self-attention mechanism block is shown in detail in Figure 1b. There are two main parts in the block: (1) the calculation of attention weight; (2) the calculation of channel feature vector with ICI.

2.2.1. The Calculation of Attention Weight

As shown in Figure 1b, the feature vectors of the i -th channel, which is the CUT, and the interfering channels, which also include the CUT itself, are used for the calculation of Q and K, respectively. For the calculation of attention weight, the two most commonly used functions are Dot-product and Additive [25]. The two functions have similar computational complexity; however, Dot-product can be implemented using highly optimized matrix multiplication, and it is much faster and more space-efficient in practice than Additive. In this work, Dot-product is chosen as the function to calculate the attention weight, and the soft-max function is applied for the normalization of attention weights. The attention weight α_{ij} can be calculated as follows:

$$\alpha_{ij} = softmax\left(F_i W^q \left(F_j W^k\right)^T\right), \tag{2}$$

where α_{ij} is the attention weight of interfering channel j for the CUT i ; $softmax(\cdot)$ is the soft-max function; W^q and W^k are the $n_d \times n_d$ (n_d is the dimension of the channel feature

vectors, $n_d = 3$ in this paper) trainable parameter matrices for the calculation of Q and K, respectively; F_i and F_j are the feature vectors of channel i and channel j , respectively.

When the attention weights of the interfering channels for the CUT i are obtained by Formula (2), the new feature vector with ICI of channel i can be calculated for the QoT estimation of channel i .

2.2.2. The Calculation of Channel Feature Vector with ICI

The attention weights of the interfering channels are the effects of the interfering channels on the CUT, and the channel feature vector with ICI is calculated as follows:

$$F'_i = \sum_j \alpha_{ij} F_j W^v, \quad (3)$$

where F'_i is the feature vector with ICI of channel i , α_{ij} is the attention weight of interfering channel j for the CUT i , F_j is the feature vector of the interfering channel j , and W^v is the $n_d \times n_d$ trainable parameter matrix for the calculation of V.

The channel feature vector with ICI F'_i contains the information of CUT i and all the interfering channels, and the channel feature vector F'_i is input to an ANN model to estimate the QoT value of the CUT i . Similarly, we can obtain the QoT values of all channels to achieve the multi-channel QoT estimation task.

3. Data Generation and Simulation Setup

In the SA-QoT-E, the input is the feature vectors of all occupied channels passing the same route, and the output is the generalized signal-to-noise ratio (GSNR) values of these channels. The GSNR can be calculated as follows:

$$GSNR = \frac{P}{P_{ASE} + P_{NLI}}, \quad (4)$$

where P is the launch power of the lightpath, P_{ASE} is the amplified spontaneous emission (ASE) noise power introduced by erbium-doped fiber amplifiers (EDFAs), and P_{NLI} is the NLI power due to the nonlinear effect of fibers. The Japan network (Jnet) and National Science Foundation network (NSFnet) are considered in our simulations, which are shown in Figure 2a,b, respectively. The transponders of the two networks are set to work on the C++ band of which the center frequency is 193.35 THz with a spectral load with 80 wavelengths (i.e., channels) on the 50 GHz spectral grid. The symbol rate of the transponders is 32 GBaud, and the launch power of each channel is uniformly selected in the range of $[-3 \sim 0]$ dBm with 0.1 dBm granularity. We assume the fibers in the two networks are ITU-T G.652 standard single-mode fibers (SSMF), of which attenuation, dispersion, and non-linearity coefficients are 0.2 dB/km, 16.7 ps/nm/km, and 1.3/W/km, respectively; the span length of the fibers is 40 km in the Jnet and 100 km in the NSFnet. The EDFAs in the two networks are set to completely compensate for fiber span losses, and the noise figure is 8 dB in the Jnet and 6.5 dB in the NSFnet. The network datasets of the two networks are generated synthetically by the open-source Gaussian noise model in the Python (GNPY) library [27], where the NLI power is calculated by the generalized Gaussian model (GNN). In each network, we generate 8000 samples for training and 2000 samples for testing by randomly choosing one of the K shortest paths of a source–destination node pair and a channel state that represents whether the channels are occupied. The training datasets in the Jnet and NSFnet are defined as D_{Jnet}^{80} and D_{NSFnet}^{80} , and the testing datasets in the Jnet and NSFnet are defined as T_{Jnet}^{80} and T_{NSFnet}^{80} . Each sample of the network dataset contains the launch power of each channel, the wavelength allocation, the transmission length, and the GSNR of each channel.

In our simulations, the performance comparison of the ANN-QoT-E and the SA-QoT-E is shown. The input of the ANN-QoT-E contains the launch power of each channel (80-dimensional vector), the wavelength allocation (80-dimensional vector), and the transmission distance, and the outputs of the ANN-QoT-E are the GSNR values of all channels (80-dimensional vector). Thus, the sizes of the first and the last layers of the ANN-QoT-E

are 161 and 80, respectively. In the ANN model of the SA-QoT-E, its input is the channel feature vector with ICI (3-dimensional vector) and its output is the GSNR value of the channel. Thus, the sizes of the first and the last layers of the ANN model of the SA-QoT-E are 3 and 1, respectively.

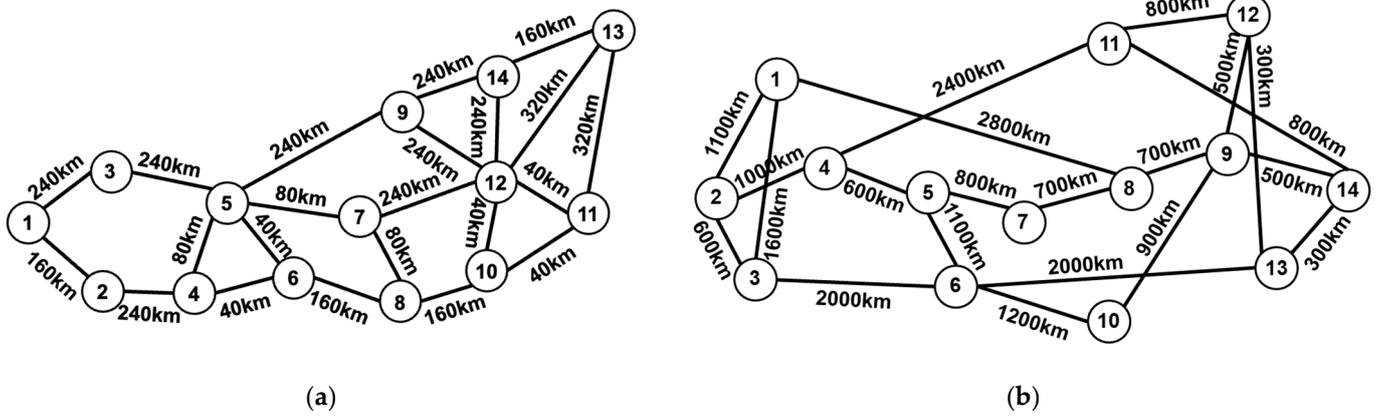


Figure 2. (a) Japan network (Jnet) topology; (b) National Science Foundation network (NSFnet) topology.

We optimize the hyperparameters of the SA-QoT-E to achieve high accuracy. To reduce the operation time of the hyperparameters search method, we empirically select the set of available hyperparameters rather than all of those. In the SA-QoT-E, we set the number of heads of the self-attention mechanism N_s , the number of hidden layers N_L , the number of neurons in hidden layers N_h , the batch size b , the learning rate α , and the epoch number e as variables with the constraints that $N_s \in \{1, 2, 3, 4\}$, $N_L \in \{1, 2, 3, 4\}$, $N_h \in \{32, 64, 128, 256, 512\}$, $b \in \{16, 32, 64, 128\}$, $\alpha \in \{0.1, 0.01, 0.001\}$, and $e \in \{100, 200, 400, 600, 800\}$. The hyperparameters of the ANN-QoT-E are similar to the ANN model of the SA-QoT-E.

After searching for the hyperparameters to achieve the highest accuracy on training datasets, the number of heads of the self-attention mechanism is set to 1 in the SA-QoT-E. The ANN-QoT-E and the ANN model of the SA-QoT-E are set to be composed of fully connected layers including 161/256/256/80 and 3/256/256/1 neurons, respectively. The activation function for all neurons is the rectified linear unit (ReLU) function and the loss function is the mean square error (MSE). In the training phase, the batch size, the learning rate, and the number of epochs of the ANN-QoT-E and the SA-QoT-E are 32, 0.01, and 400, respectively. We extract 1/10 of the training set as the verification set, and the trained mode is verified every 50 epochs and the current optimal model is saved.

4. Simulation Results and Analyses

Figure 3a,b show the GSNR decrease of the Ch_1, Ch_20, Ch_40, Ch_60, and Ch_80 caused by the deployment of the new lightpaths in the Jnet and the NSFnet. In the Jnet and the NSFnet, the selected routes are 1-3-5-9-14-13 and 2-4-11-12, respectively, and the wavelength assignment scheme is the first-fit (FF). In Figure 3a, with the increase in the number of newly deployed lightpaths, the GSNR decrease of the CUTs increases and the maximum GSNR decrease achieves 3.25 dB. Moreover, due to the FF scheme, the GSNR of the channel with the smallest number (Ch_1) decreases the most when the number of newly deployed lightpaths is 10, and when the number of newly deployed lightpaths achieves 70, the GSNR of the channel with the middle number (Ch_40) decreases the most. In Figure 3b, the performance of the GSNR decrease in the NSFnet is similar to that in the Jnet, and the maximal GSNR decrease of the CUTs is 1.36 dB. The QoT of the previously deployed lightpaths is deteriorated greatly due to the deployment of new lightpaths, and the QoT estimation of the single lightpath cannot capture the decrease in the previously

deployed lightpaths' QoT. Thus, it is necessary to estimate the QoT of new-LPs and pre-LPs simultaneously, i.e., achieve multi-channel QoT estimation, in optical networks.

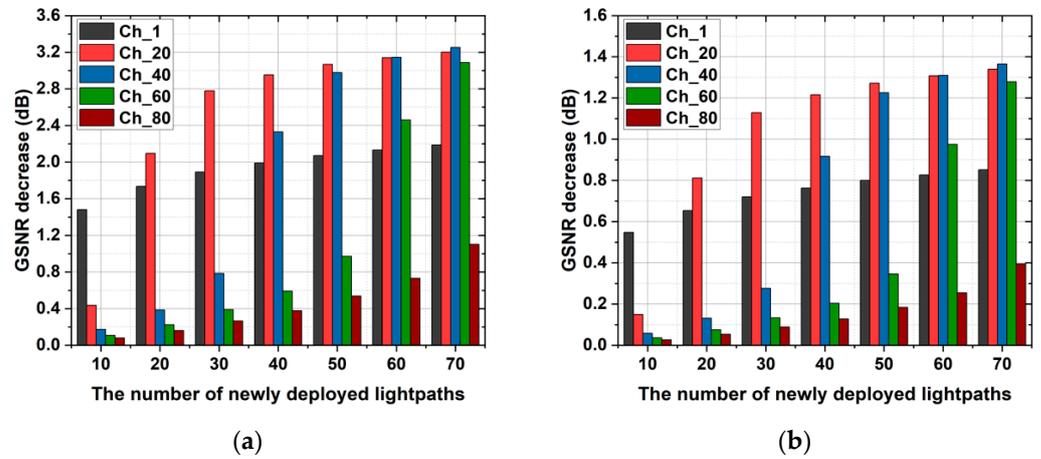


Figure 3. (a) GSNR decrease in the Jnet; (b) GSNR decrease in the NSFnet.

In this section, we show the performance of the ANN-QoT-E proposed in [23] and our proposed SA-QoT-E in respect of model convergence, estimation accuracy, scalability, and computational complexity.

4.1. Convergence

The training processes of the ANN-QoT-E and the SA-QoT-E in the Jnet and NSFnet are shown in Figure 4a,b. As shown in Figure 4a, in the Jnet, the ANN-QoT-E and the SA-QoT-E converge after 400 epochs. After 79 epochs, the ANN-QoT-E almost converges and its training loss is 0.127, and the SA-QoT-E almost converges at the 51th epoch and its training loss is 0.127. In Figure 4b, after 400 epochs, the two models converge in the NSFnet, where the SA-QoT-E almost converges at the 53th epoch with the training loss of 0.059 and the ANN-QoT-E almost converges at 226th epoch with the training loss of 0.057. In conclusion, the SA-QoT-E converges in fewer epochs than the ANN-QoT-E. That is because the SA-QoT-E has a stronger data-fitting ability than the ANN-QoT-E.

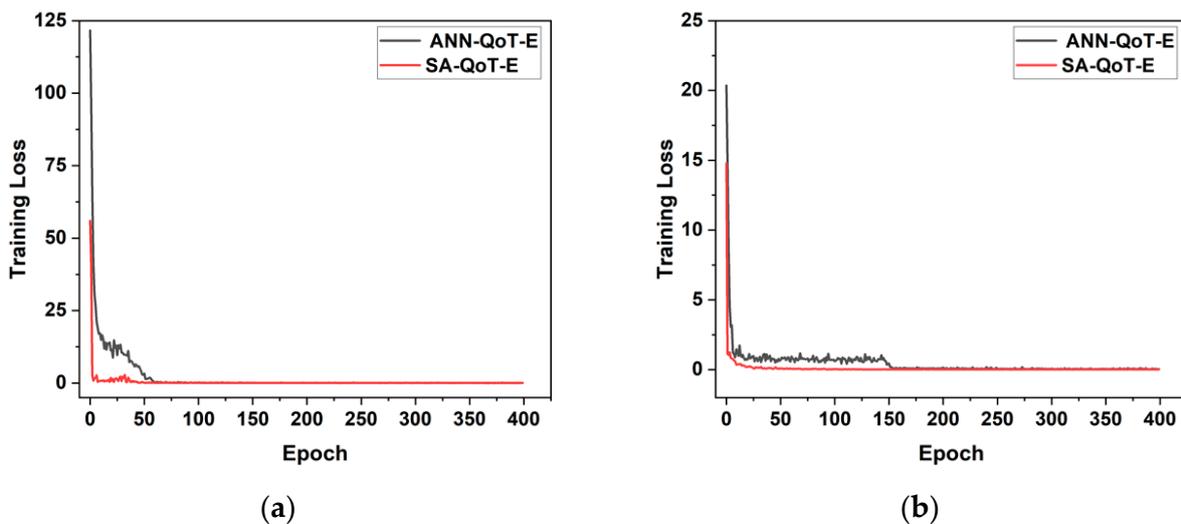


Figure 4. (a) Training loss vs. the number of epochs for the ANN-QoT-E and the SA-QoT-E in the Jnet; (b) training loss vs. the number of epochs for the ANN-QoT-E and the SA-QoT-E in the NSFnet.

4.2. Estimation Accuracy

Figure 5 shows the estimation accuracy of the ANN-QoT-E and the SA-QoT-E tested on the testing dataset. Figure 5a shows the testing of the mean absolute error (MAE) of the two estimation models in the Jnet and NSFnet. The testing MAE of the SA-QoT-E is lower than that of the ANN-QoT-E; specifically, the testing MAE of the SA-QoT-E is 0.147 dB and 0.127 dB lower than that of the ANN-QoT-E in the Jnet and NSFnet, respectively, which means the predicted GSNR of the SA-QoT-E is, on average, 0.147 dB and 0.127 dB closer to the ground truth of the GSNR compared with that of the ANN-QoT-E in the Jnet and NSFnet, respectively. R2 score is a commonly used metric for the evaluation of the regression model, it is closer to 1, and the corresponding model has higher estimation accuracy. Figure 5b shows that, in the Jnet and the NSFnet, the R2 score of the SA-QoT-E is 0.03 and 0.016 higher than that of the ANN-QoT-E, respectively.

Figure 5c,d show the predicted GSNR of the ANN-QoT-E and the SA-QoT-E against their actual GSNR in the Jnet, respectively. The ideal estimation result is that the scatters distribute on the baseline, which means the predicted GSNR is equal to the actual GSNR. The figures show that, in the Jnet, the estimation accuracy of the SA-QoT-E is higher than that of the ANN-QoT-E; specifically, the maximum absolute error of the SA-QoT-E is 1.13 dB and that of the ANN-QoT-E is 4.21 dB.

Similarly, Figure 5e,f show the predicted GSNR values of the two models against their actual GSNR in the NSFnet. The results show that the estimation accuracy of the SA-QoT-E is higher than that of the ANN-QoT-E in the NSFnet, where the maximum absolute error of the SA-QoT-E is 0.84 dB and that of the ANN-QoT-E is 3.46 dB.

In the Jnet and the NSFnet, the accuracy of the SA-QoT-E is higher than that of the ANN-QoT-E, which is because the self-attention mechanism assigns weights to the interfering channels according to their effects on the QoT of CUT. Figure 6a,b show the attention weight and launch power of each channel for Ch_16 in the Jnet and for Ch_76 in the NSFnet, respectively, which is tested by the SA-QoT-E on a randomly selected testing sample. Due to the nonlinear effect of fibers, the interfering channel, which has a higher launch power and closer spectral distance to the CUT (SDC), has a stronger effect on the CUT. Figure 6a marks the channels with the larger attention weight. In the Jnet, as shown in Figure 6a, these marked channels are arranged in descending order of attention weight as Ch_16, Ch_7, Ch_14, Ch_75, Ch_72, Ch_13, Ch_60, and Ch_25. The attention weight of Ch_16 is maximal due to its launch power being maximal and SDC being minimal. Though the SDC of Ch_7 is higher than that of Ch_14, the attention weight of Ch_7 is higher than that of Ch_14 due to the higher launch power of Ch_7. The attention weights assigned for Ch_13, Ch_25, Ch_60, Ch_72, and Ch_75 violate the nonlinear effect theory; thus, the accuracy of the SA-QoT-E in the Jnet is lower than that in the NSFnet. Figure 6b shows that, in the NSFnet, the order of the marked channels according to their attention weight is Ch_76, Ch_29, Ch_10, Ch_69, and Ch_41. The attention weight of Ch_76 is maximal due to the maximal launch power and minimal SDC of Ch_76, and the attention weight of Ch_29 and Ch_10 is higher than that of Ch_69 and Ch_41, which is because the launch power of Ch_29 and Ch_10 is higher than that of Ch_69 and Ch_41. Due to the smaller SDC of Ch_29 than Ch_10, the attention weight of Ch_29 is higher than that of Ch_10. For the same reason, the attention weight of Ch_69 is higher than that of Ch_41. The attention weights assignment in the NSFnet obeys the nonlinear effect theory.

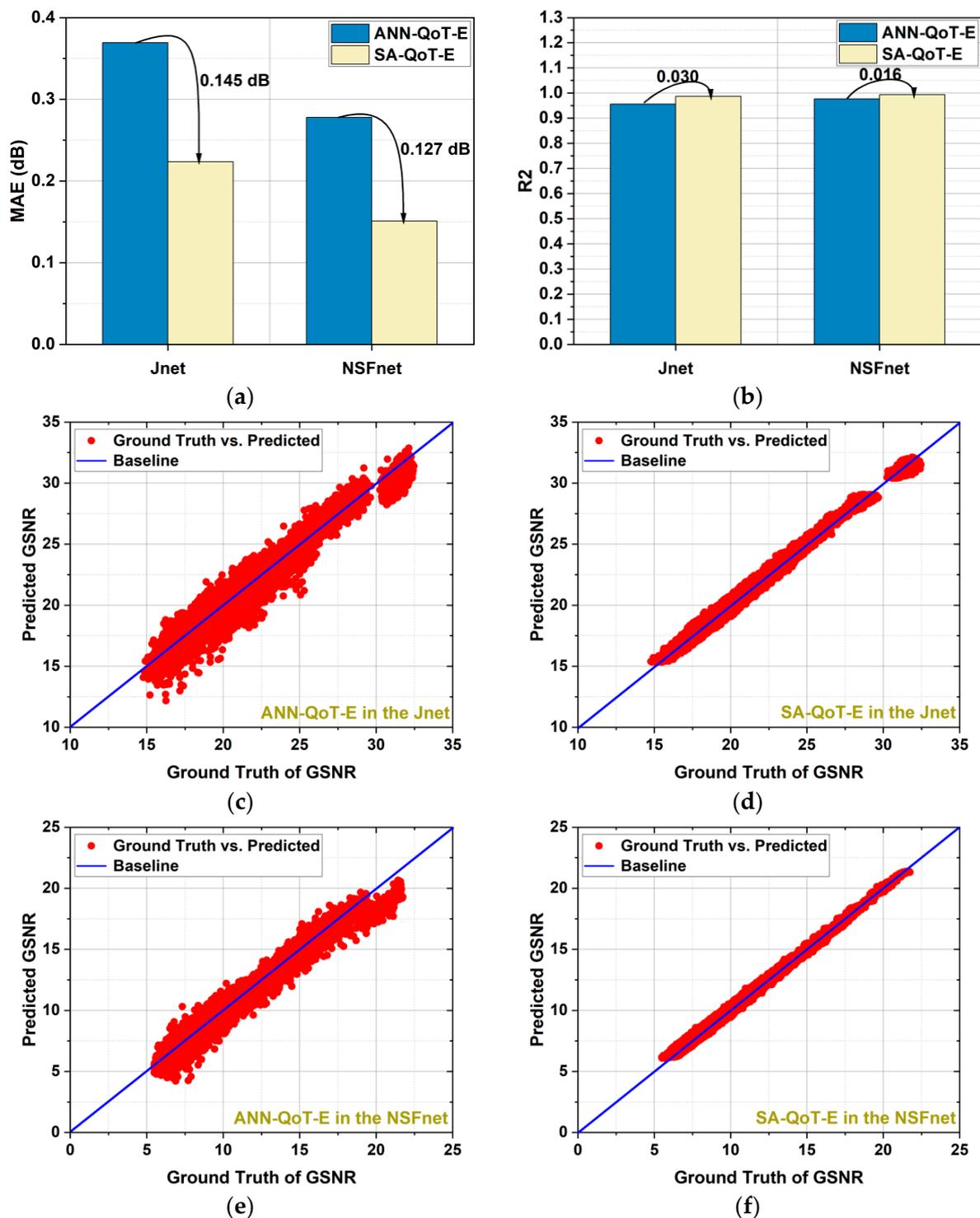


Figure 5. Estimation accuracy of the ANN-QoT-E and the SA-QoT-E: (a) testing MAE of two models in the Jnet and NSFnet; (b) testing R2 of two models in the Jnet and NSFnet; (c) ground truth of GSNR vs. predicted GSNR based on the ANN-QoT-E in the Jnet; (d) ground truth of GSNR vs. predicted GSNR based on the SA-QoT-E in the Jnet; (e) ground truth of GSNR vs. predicted GSNR based on the ANN-QoT-E in the NSFnet; (f) ground truth of GSNR vs. predicted GSNR based on the SA-QoT-E in the NSFnet.

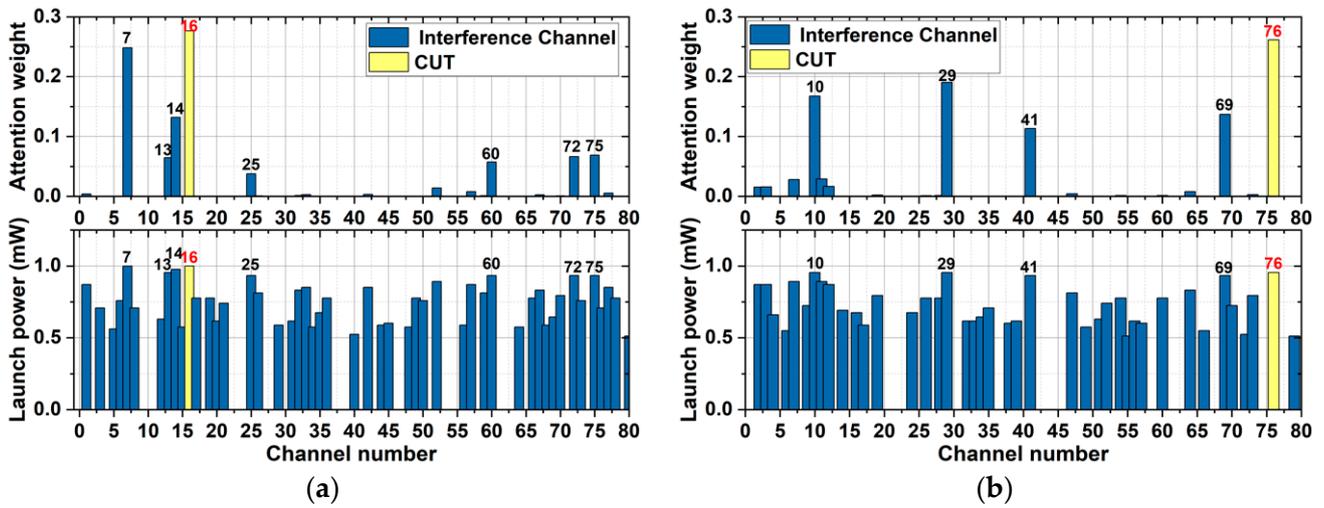


Figure 6. (a) Attention weight and launch power of each channel for Ch_16 in the Jnet; (b) attention weight and launch power of each channel for Ch_76 in the NSFnet.

4.3. Scalability

The input dimension and output dimension of ANN models are fixed. When the number of network wavelengths is expanded (such as a C band network expanding to a C+L band network), the original ANN-QoT-E cannot be applied to the network. The SA-QoT-E has the advantage of variable length of the input feature vector sequence due to the introduction of the self-attention mechanism. We generate new testing datasets in the Jnet with 120 wavelengths (the full channels of the C++ band), in the Jnet with 216 wavelengths (the full channels of the C+L band), in the NSFnet with 120 wavelengths, and in the NSFnet with 216 wavelengths, which are defined as T_{Jnet}^{120} , T_{Jnet}^{216} , T_{NSFnet}^{120} and T_{NSFnet}^{216} , respectively. Figure 7 shows the estimation accuracy of the SA-QoT-E in the Jnet and NSFnet obtained by testing on $T_{Jnet}^{80}/T_{Jnet}^{120}/T_{Jnet}^{216}$ and $T_{NSFnet}^{80}/T_{NSFnet}^{120}/T_{NSFnet}^{216}$, where the SA-QoT-E is trained on $D_{Jnet}^{80}/D_{NSFnet}^{80}$. As shown in Figure 7a, in the Jnet, the testing MAEs of the SA-QoT-E tested on $T_{Jnet}^{80}/T_{Jnet}^{120}/T_{Jnet}^{216}$ are low; even the highest MAE obtained on T_{Jnet}^{216} is lower than the MAE obtained by the ANN-QoT-E in Figure 5a; in the NSFnet, the testing MAEs of the SA-QoT-E tested on $T_{NSFnet}^{80}/T_{NSFnet}^{120}/T_{NSFnet}^{216}$ are close and low. Figure 7b shows the R2 score of the SA-QoT-E tested on $T_{Jnet}^{80}/T_{Jnet}^{120}/T_{Jnet}^{216}$ and $T_{NSFnet}^{80}/T_{NSFnet}^{120}/T_{NSFnet}^{216}$ in the Jnet and NSFnet, and the SA-QoT-E achieves a high R2 score.

Figure 7c,d show the predicted value of the SA-QoT-E against their actual GSNR tested on T_{Jnet}^{120} and T_{Jnet}^{216} , respectively. The results show that the SA-QoT-E trained on D_{Jnet}^{80} has a relatively high accuracy on T_{Jnet}^{120} and T_{Jnet}^{216} ; specifically, the maximum absolute errors of the SA-QoT-E tested on T_{Jnet}^{120} and T_{Jnet}^{216} are 1.79 dB and 2.58 dB, respectively.

Figure 7e,f show that the SA-QoT-E trained on D_{NSFnet}^{80} has a relatively high accuracy on T_{NSFnet}^{120} and T_{NSFnet}^{216} ; specifically, the maximum absolute errors of the SA-QoT-E tested on T_{NSFnet}^{120} and T_{NSFnet}^{216} are 0.84 dB and 0.73 dB, respectively.

The self-attention mechanism learns the interaction between channels, and aggregates multiple channel feature vectors into a new channel feature vector to estimate the QoT of the corresponding channel, Thus, the SA-QoT-E is not limited by the number of network channels. In conclusion, the SA-QoT-E can be directly applied to network wavelength expansion scenarios without retraining.

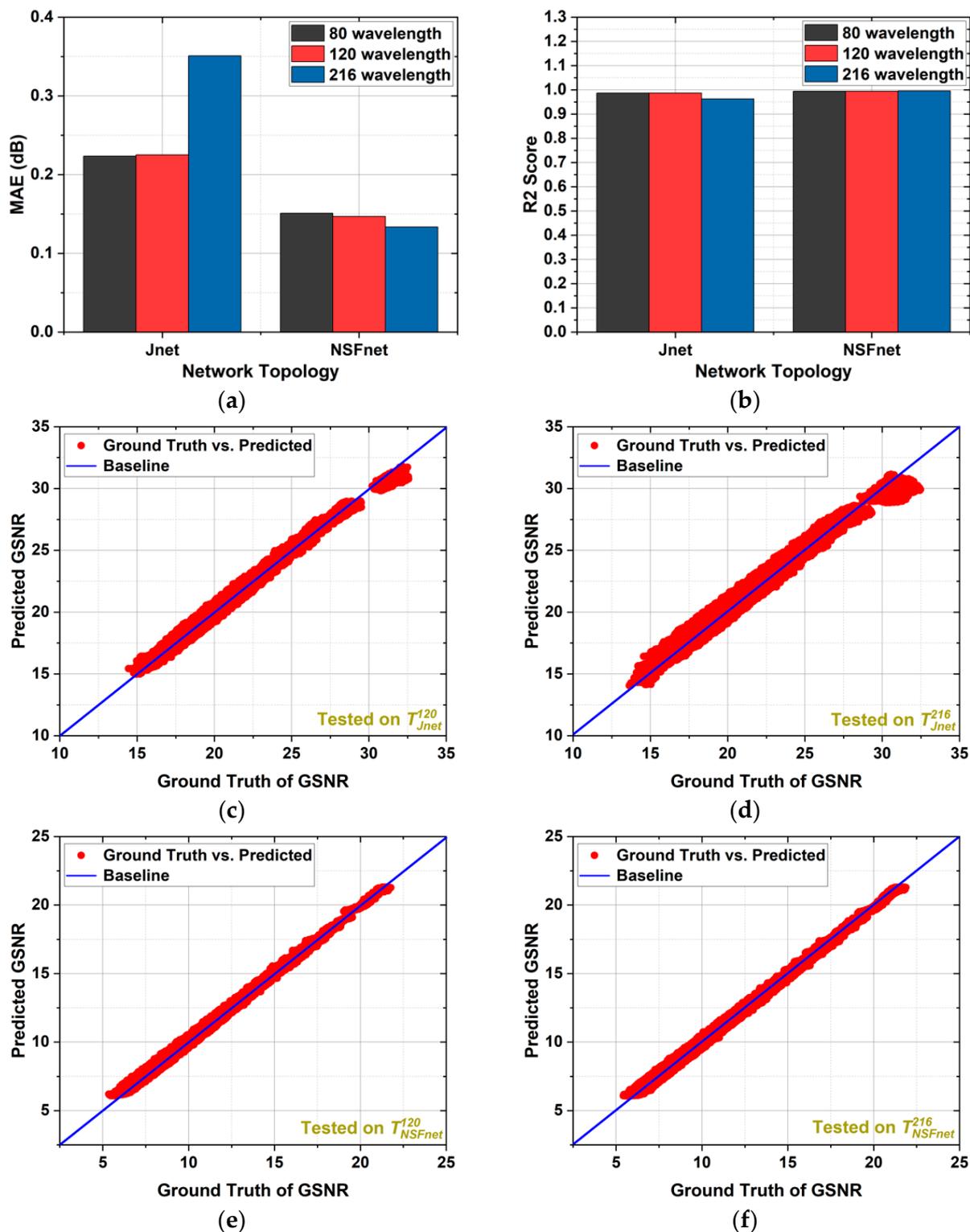


Figure 7. Estimation accuracy of the SA-QoT-E in the Jnet and NSFnet with different numbers of wavelengths: (a) testing MAE of the SA-QoT-E in the Jnet and NSFnet; (b) testing R2 of the SA-QoT-E in the Jnet and NSFnet; (c) ground truth of GSNR vs. predicted GSNR based on the SA-QoT-E in the Jnet with 120 wavelengths; (d) ground truth of GSNR vs. predicted GSNR based on the SA-QoT-E in the Jnet with 216 wavelengths; (e) ground truth of GSNR vs. predicted GSNR based on the SA-QoT-E in the NSFnet with 120 wavelengths; (f) ground truth of GSNR vs. predicted GSNR based on the SA-QoT-E in the NSFnet with 216 wavelengths.

4.4. Computational Complexity

Compared with the ANN-QoT-E, the SA-QoT-E not only improves the estimation accuracy but also has scalability. However, the SA-QoT-E has higher computational complexity. The computational complexity is the total number of addition operations and multiplication operations of the QoT estimators. The computational complexity of each layer of the ANN models is its input dimension multiplied by its output dimension; thus, the computational complexity of the ANN-QoT-E in our simulations can be calculated as follows:

$$C_{ANN} = O\left((2N_{ch} + 1)N_h + N_h^2 + N_h N_{ch}\right) = O\left(N_{ch}N_h + N_h^2\right), \quad (5)$$

where C_{ANN} is the computational complexity of the ANN-QoT-E, N_{ch} is the number of wavelengths in the network, and N_h is the number of neurons in the ANN model's hidden layer.

The self-attention mechanism mainly contains three steps: (1) similarity calculation; (2) soft-max operation; (3) weighted summation. First, the computational complexity of the similarity calculation is $O(N_{ch}^2 n_d)$, due to the fact that it is a $N_{ch} \times n_d$ matrix multiplied by a $n_d \times N_{ch}$ matrix. In addition, the computational complexity of soft-max operation is $O(N_{ch}^2)$. Finally, the weighted summation is a $N_{ch} \times N_{ch}$ matrix multiplied by a $N_{ch} \times n_d$ matrix; thus, its computational complexity is $O(N_{ch}^2 n_d)$. The computational complexity of the ANN model in the SA-QoT-E is $O(N_{ch}(n_d N_h + N_h^2 + N_h))$. Therefore, the computational complexity of the SA-QoT-E is shown as follows:

$$C_{SA} = O\left(2N_{ch}^2 n_d + N_{ch}^2 + N_{ch}(n_d N_h + N_h^2 + N_h)\right) = O\left(N_{ch}^2 n_d + N_{ch} N_h n_d + N_{ch} N_h^2\right), \quad (6)$$

where C_{SA} is the computational complexity of the SA-QoT-E. The computational complexity of the ANN-QoT-E and the SA-QoT-E is shown in Table 2. Obviously, the computational complexity of the SA-QoT-E is higher than that of the ANN-QoT-E. However, the training of the estimation models is offline and C_{SA} is acceptable, and the SA-QoT-E can be applied to realistic optical network optimization.

Table 2. Computational complexity of the ANN-QoT-E and the SA-QoT-E.

QoT Estimation Model	Computational Complexity
ANN-QoT-E	$O(N_{ch}N_h + N_h^2)$
SA-QoT-E	$O(N_{ch}^2 n_d + N_{ch}N_h n_d + N_{ch}N_h^2)$

4.5. Discussion

The estimation accuracy of QoT estimators is important to ensure the reliable transmission of the lightpaths in optical networks, and the computational complexity of QoT estimators decides their availability in practical scenarios. The SA-QoT-E has the advantages compared with the ANN-QoT-E: (1) stronger data-fitting ability; (2) higher estimation accuracy; (3) stronger scalability. However, these advantages of SA-QoT-E come at the cost of high computational complexity. The computational complexity of SA-QoT-E is higher than that of the ANN-QoT-E, which is mainly affected by the number of channels N_{ch} . Thus, the ANN-QoT-E is more suitable for optical networks with a large number of network channels and a shortage of computing resources. In most scenarios, the SA-QoT-E has more advantages compared with the ANN-QoT-E.

5. Conclusions

In multi-channel optical networks, due to the nonlinear effect of fibers, the different interfering channel has a different effect on the CUT. The existing ANN-QoT-E ignores the different effects of the interfering channels, which affects its estimation accuracy. To improve the accuracy of the ANN-QoT-E, we propose a novel SA-QoT-E, where the self-attention mechanism assigns attention weights to the interfering channels according to

their effects on the CUT. Moreover, we use a hyperparameter search method to optimize the hyperparameters of the SA-QoT-E. The simulation results show that the proposed SA-QoT-E improves the estimation accuracy compared with the ANN-QoT-E. Specifically, compared with the ANN-QoT-E, the testing MAE achieved by the SA-QoT-E is decreased by 0.147 dB and 0.127 dB in the Jnet and NSFnet, respectively; the R2 score achieved by the SA-QoT-E is improved by 0.03 and 0.016 in the Jnet and NSFnet, respectively; the maximal absolute error achieved by the SA-QoT-E is reduced from 4.21 dB to 1.13 dB in the Jnet and from 3.46 dB to 0.84 dB in the NSFnet. Moreover, the SA-QoT-E has scalability, which can be directly applied to network wavelength expansion scenarios without retraining. However, compared with the ANN-QoT-E, the SA-QoT-E has higher computational complexity. Fortunately, the training of the SA-QoT-E is offline and the computational complexity of the SA-QoT-E is acceptable; thus, the proposed SA-QoT-E can be applied to the realistic optical network and provide more accurate lightpath QoT information for optical network optimization.

Author Contributions: Conceptualization, Y.Z. and Z.G.; methodology, Y.Z. and Z.G.; software, Y.Z. and Z.G.; validation, Y.Z., X.H. and Z.G.; formal analysis, J.Z. and Y.D.; investigation, J.Z. and R.G.; resources, J.Z. and Y.J.; data curation, J.Z. and Z.G.; writing—original draft preparation, Y.Z. and Z.G.; writing—review and editing, Z.G., J.Z., R.G. and Y.J.; visualization, Y.Z.; supervision, X.H. and Y.D.; project administration, X.H. and Y.D.; funding acquisition, Z.G. and R.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (No. 62101058), and in part by the Fund of State Key Laboratory of IPOC (BUPT) (No. IPOC2022ZT11), P.R. China.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ayassi, R.; Triki, A.; Crespi, N.; Minerva, R.; Laye, M. Survey on the use of machine learning for quality of transmission estimation in optical transport networks. *J. Light. Technol.* **2022**, *40*, 5803–5815. [[CrossRef](#)]
2. Shao, J.; Liang, X.; Kumar, S. Comparison of Split-Step Fourier Schemes for Simulating Fiber Optic Communication Systems. *IEEE Photonics J.* **2014**, *6*, 7200515.
3. Poggiolini, P. The GN Model of Non-Linear Propagation in Uncompensated Coherent Optical Systems. *J. Light. Technol.* **2012**, *30*, 3857–3879. [[CrossRef](#)]
4. Aladin, S.; Tremblay, C. Cognitive Tool for Estimating the QoT of New Lightpaths. In Proceedings of the 2018 Optical Fiber Communications Conference and Exposition (OFC), San Diego, CA, USA, 11–15 March 2018.
5. Rottondi, C.; Barletta, L.; Giusti, A.; Tornatore, M. Machine-learning method for quality of transmission prediction of unestablished lightpaths. *IEEE/OSA J. Opt. Commun. Netw.* **2018**, *10*, A286–A297. [[CrossRef](#)]
6. Azzimonti, D.; Rottondi, C.; Tornatore, M. Reducing probes for quality of transmission estimation in optical networks with active learning. *IEEE/OSA J. Opt. Commun. Netw.* **2020**, *12*, A38–A48. [[CrossRef](#)]
7. Rottondi, C.; Riccardo, D.M.; Mirko, N.; Alessandro, G.; Andrea, B. On the benefits of domain adaptation techniques for quality of transmission estimation in optical networks. *IEEE/OSA J. Opt. Commun. Netw.* **2021**, *13*, A34–A43. [[CrossRef](#)]
8. Liu, C.-Y.; Chen, X.; Proietti, R.; Yoo, S.J.B. Performance studies of evolutionary transfer learning for end-to-end QoT estimation in multi-domain optical networks. *IEEE/OSA J. Opt. Commun. Netw.* **2021**, *13*, B1–B11. [[CrossRef](#)]
9. Samadi, P.; Amar, D.; Lepers, C.; Lourdiane, M.; Bergman, K. Quality of transmission prediction with machine learning for dynamic operation of optical WDM networks. In Proceedings of the 2017 European Conference on Optical Communication (ECOC), Gothenburg, Sweden, 17–21 September 2017.
10. Seve, E.; Pesic, J.; Delezoide, C.; Bigo, S.; Pointurier, Y. Learning Process for Reducing Uncertainties on Network Parameters and Design Margins. *IEEE/OSA J. Opt. Commun. Netw.* **2018**, *10*, A298–A306. [[CrossRef](#)]
11. Yu, J.; Mo, W.; Huang, Y.K.; Lp, E.; Kilper, D.C. Model transfer of QoT prediction in optical networks based on artificial neural networks. *IEEE/OSA J. Opt. Commun. Netw.* **2019**, *11*, C48–C57. [[CrossRef](#)]
12. Mahajan, A.; Christodoulopoulos, K.; Martinez, R.; Spadaro, S.; Muñoz, R. Modeling EDFA gain ripple and filter penalties with machine learning for accurate QoT estimation. *J. Light. Technol.* **2020**, *38*, 2616–2629. [[CrossRef](#)]

13. Cho, H.J.; Varughese, S.; Lippiatt, D.; Ralph, S.E. Convolutional recurrent machine learning for OSNR and launch power estimation: A critical assessment. In Proceedings of the 2020 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 28 March–1 April 2020.
14. Azzimonti, D.; Rottondi, C.; Giusti, A.; Tornatore, M.; Bianco, A. Comparison of domain adaptation and active learning techniques for quality of transmission estimation with small-sized training datasets. *IEEE/OSA J. Opt. Commun. Netw.* **2021**, *13*, A56–A66. [[CrossRef](#)]
15. Liu, X.; Lun, H.; Liu, L.; Zhang, Y.; Liu, Y.; Yi, L.; Hu, W.; Zhuge, Q. A Meta-Learning-Assisted Training Framework for Physical Layer Modeling in Optical Networks. *J. Light. Technol.* **2022**, *40*, 2684–2695. [[CrossRef](#)]
16. Kruse, L.E.; Kühn, S.; Pachnicke, S. Exact component parameter agnostic QoT estimation using spectral data-driven LSTM in optical networks. In Proceedings of the 2022 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 7–11 March 2022.
17. Bergk, G.; Shariati, B.; Safari, P.; Fischer, J.K. ML-assisted QoT estimation: A dataset collection and data visualization for dataset quality evaluation. *IEEE/OSA J. Opt. Commun. Netw.* **2022**, *14*, 43–55. [[CrossRef](#)]
18. Ayoub, O.; Bianco, A.; Andreoletti, D.; Troia, S.; Giordano, S.; Rottondi, C. On the Application of Explainable Artificial Intelligence to Lightpath QoT Estimation. In Proceedings of the 2022 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 5–9 March 2022.
19. Ayoub, O.; Andreoletti, D.; Troia, S.; Giordano, S.; Bianco, A.; Rottondi, C. Quantifying Features' Contribution for ML-based Quality-of-Transmission Estimation using Explainable AI. In Proceedings of the 2022 European Conference on Optical Communication (ECOC), Basel, Switzerland, 18–22 September 2022.
20. Allogba, S.; Aladin, S.; Tremblay, C. Machine-learning-based lightpath QoT estimation and forecasting. *J. Light. Technol.* **2022**, *40*, 3115–3127. [[CrossRef](#)]
21. Panayiotou, T.; Savva, G.; Shariati, B.; Tomkos, I.; Ellinas, G. Machine Learning for QoT Estimation of Unseen Optical Network States. In Proceedings of the 2019 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 3–7 March 2019.
22. Safari, P.; Shariati, B.; Bergk, G.; Fischer, J.K. Deep Convolutional Neural Network for Network-wide QoT Estimation. In Proceedings of the 2021 Optical Fiber Communications Conference and Exhibition (OFC), San Francisco, CA, USA, 6–10 June 2021.
23. Gao, Z.; Yan, S.; Zhang, J.; Mascarenhas, M.; Nejabati, R.; Ji, Y.; Simeonidou, D. ANN-based multi-channel QoT-prediction over a 563.4-km field-trial testbed. *J. Light. Technol.* **2020**, *38*, 2646–2655. [[CrossRef](#)]
24. Zhou, Y.; Gu, Z.; Zhang, J.; Ji, Y. Attention Mechanism Based Multi-Channel QoT Estimation in Optical Networks. In Proceedings of the 2021 Asia Communications and Photonics Conference (ACP), Shanghai, China, 24–27 October 2021.
25. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
26. Jain, A.K.; Mao, J.; Mohiuddin, K.M. Artificial neural networks: A tutorial. *Computer* **1996**, *29*, 31–44. [[CrossRef](#)]
27. Ferrari, A.; Filer, M.; Balasubramanian, K.; Yin, Y.; Le Rouzic, E.; Kundrat, J.; Grammel, G.; Galimberti, G.; Curri, V. GNPpy: An open source application for physical layer aware open optical networks. *IEEE/OSA J. Opt. Commun. Netw.* **2020**, *12*, C31–C40. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.