Multi-task optical performance monitoring using a transfer learning assisted cascaded deep neural network in WDM systems

Yameng Cao,[1] Di Zhang, [1,2] Hanyu Zhang,[1] and Yan Ling Xue[1,2,3,4*]
[1] Shanghai Key Laboratory of Multidimensional Information Processing, School of Communication and Electronics Engineering, East China Normal University, Shanghai 200241, China
[2] Shanghai Branch, Hefei National Laboratory, Shanghai 201315, China
[3] Shanghai Research Center for Quantum Sciences, Shanghai 201315, China
[4] Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China
* ylxue@ee.ecnu.edu.cn

***Abstract*—We design a transfer learning assisted cascaded deep neural network (TL-CDNN) for multi-task optical performance monitoring in a WDM system. The method can decrease training resource and is robust to interference in WDM systems.**

Keywords—optical performance monitoring, modulation format identification, OSNR estimation, wavelength division multiplexing, cascaded deep neural network, transfer learning

I. Introduction

With the explosive growth of transmitted data, the next generation optical networks should be with more flexibility in managing the most of system parameters. The optical performance monitoring (OPM) technology, which can measure the critical system parameters such as optical signal-to-noise ratio (OSNR), bit rate and modulation format etc., is one of the keys to ensure the reliability of the next generation networks.

The wavelength division multiplexing (WDM) technology is the most widely used capacity expansion method. However, the OPM schemes for WDM systems have been less studied. If the current ordinary OPM schemes for a single channel are directly applied to a WDM system with multiple channels, the analytical model training processes have to be repeated for each channel individually. It would be excessively time-consuming.

Over the past decade, deep learning has attracted the attention of researchers in the OPM field which can be used for the accurate monitoring of a certain key parameter [1-3] or multiple parameters simultaneously [4-6]. It has also been combined with various neural network architectures for the OPM tasks in different transmission scenarios for the enhancement in efficiency and performance [7-9].

Transfer learning (TL) has become a hot spot in deep learning because of its ability to apply knowledge and skills learned in previous tasks to novel tasks [10]. So once the link conditions change, transfer learning can greatly reduce the required amount of training samples and re-training time. Existing works on TL deployment in OPM tend to focus on analyzing the performance of conventional convolutional neural networks, and transfer the prior knowledge learned from traditional computer vision database ImageNet to various propagating modes [11], or building models on real-time variable WDM channels to complete the migration between different modulation formats and transmission rates [12]. However, the generalization of pre-trained models based on single-channel and the robustness of fine-tuned models based on complex WDM channels have not been studied in details.

The similarity between single-channel systems and WDM systems reminds us the possibility to apply transfer learning in a WDM system. To the best of our knowledge, we are the first so far to apply transfer learning from a single-channel system to a multi-channel WDM system. We use the data collected in the single-channel system as the source domain, train the model in advance, and then use the data of each channel in the WDM system to fine-tune model parameters and complete the OPM task of each channel.

In this paper, we propose a transfer learning assisted cascaded deep neural network (TL-CDNN) for joint modulation format identification (MFI)-OSNR monitoring task in a three-channel WDM system. We deal with four modulation formats: 8QAM, 16QAM, 32QAM and 64QAM. Considering the aging of devices and the system flexibility, we also deal with the situation of different signal power in three channels. The results demonstrate that the design can greatly decrease the training dataset size while maintaining good performance and exhibits strong robustness to the interference caused by the crosstalk between adjacent channels which is unique in WDM systems.

II. Operating Principle

In this section, we will discuss details about the extraction of selected signal features and TL-CDNN model used in our monitoring task. After simple compensation for linear impairments by modulation format independent algorithms, the signals at the receiver will be used to extract data features. Due to the significant impact on the shape of amplitude histograms (AHs) brought by different modulation formats and OSNR values, AHs are suitable to be chosen as the input features of both MFI and OSNR estimation tasks.

A deep neural network is composed of an input layer, hidden layers and an output layer. The neurons of each layer are fully connected with the neurons in last layer. Through multi-layer feature extraction and mapping, the neural network can fit highly complex nonlinear relations. The output of the k-th layer neurons can be written as:

$$Y_k = f_k(W_k \cdot Y_{k-1}),\qquad (1)$$

where $Y_k \in R^m$ represents a vector consisting of the values of m neurons in the k-th layer; $Y_{k-1} \in R^n$ represents a vector consisting of the values of n neurons in the last layer. $W_k \in R^{m\times n}$ is a weight matrix with the shape $m\times n$; f_k represents the activation function used at the k-th layer.

For the MFI task, the parameter to be observed are discrete, corresponding to the classification task in deep learning with cross-entropy [13] as the loss function. For the OSNR estimation task, the parameters to be observed are continuous, corresponding to the regression task in deep learning with mean square error (MSE) [14] as the loss function.

Previously, different networks have been used for OSNR estimation or MFI individually. However, in this paper, we have designed a cascaded deep neural network (CDNN) for joint OSNR estimation and MFI. As shown in Fig. 1, the CDNN consists of a classification network followed by a regression network, and the amplitude histogram data is the common input of these two-level networks. The first-level network is used to identify the modulation format of the input sample and then automatically selects the second-level network to estimate the OSNR value according to the output label. In this paper, each amplitude histogram is divided into 80 bins, so the number of neurons in the input layer of two networks is 80. The composition of hidden layers used to extract features in the two network is the same, the number of neurons in the hidden layers is 200 and 100, respectively, and the nonlinear activation function used by each neuron is LeakyReLU. Moreover, since the output of neurons is naturally continuous and interval free, we need to use Softmax function as the activation function of the output layer of the classification model to transform the output value into a multi-dimensional vector with continuous values between [0, 1], representing the probability that the input sample belongs to various types. In the output layer, the number of neurons in the MFI model is 4, corresponding to four modulation formats. The number of neurons in the OSNR estimation model is 1, representing the OSNR value.

However, in existing OPM techniques, training the optimal parameters of a deep learning model requires large numbers of training samples with same distribution as the test set. Once the settings of the transmission link change, it needs to be retrained. In a WDM system, due to the inevitability of inter-channel interference, distinctions exist between different wavelength channels. If multiple channels are monitored simultaneously, the model training should be performed for each individual channel. This costs training samples and training time multiplied, which is quite resource-consuming.

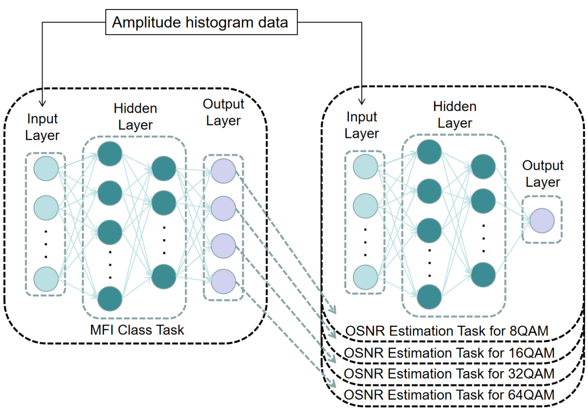


Fig. 1. CDNN structure with AH data as input and the result of class task will be used to select OSNR estimation module.

To solve this problem, we introduce transfer learning. Taking the data collected in a simple single-channel system as the source domain, only a small amount of data from different WDM channels in the target domain need to be collected. Transfer learning can reduce the dataset size while ensuring the OPM performance by obtaining relevant knowledge from the source domain.

Besides, there are two situations resulting in the deterioration of transmission signal quality: (a) the decrease of the signal power in the channel under test, caused by system aging or temperature changing; and (b) the aggravated interference from the adjacent channel caused by the higher signal power in the adjacent channel due to the flexibility in signal power across different WDM channels in future optical networks. These two situations are both taken into account to verify the performance of the proposed OPM method in the WDM system in this work.

III. System Setup and Results

A. System setup

Based on the simulation software OptiSystem, we use a pseudo-random binary sequence (PRBS) to digitally simulate 8QAM, 16QAM, 32QAM and 64QAM with the sequence length of $2^{15}-1$. The self-phase modulation and cross-phase modulation are introduced into the transmission, then the modulated signals will be launched to a span of single-mode fiber (SMF) whose transmission length is $3*80$ km, dispersion parameter is 16.75 ps/nm/km, nonlinear coefficient is 2.6×10^{-20} m^2/W and attenuation coefficient is 0.2 dB/km. The generated signals are transmitted in a single-channel system and a three-channel WDM system

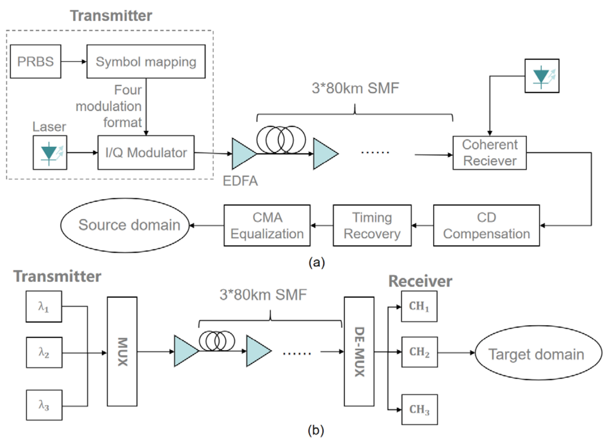


Fig. 2. (a)Schematic diagram of the single-channel system; (b)schematic diagram of the three-channel WDM system (EDFA: Erbium-doped fiber amplifier; MUX: multiplexer; DE-MUX: demultiplexer).

respectively. The data collected from the single-channel system is used as the source domain, as shown in Fig. 2(a). The data collected from the WDM system is used as the target domain, as shown in Fig. 2(b).

At the transmitting side of the single-channel system, the central wavelength of the laser is 193.4 THz and the line width is 100 kHz. At the transmitting side of the three-channel WDM system, the central wavelengths of the three lasers used are 193.3 THz, 193.4 THz and 193.5 THz, respectively, and the line width is also 100 kHz. The channel at 193.4 THz in the WDM system is selected as the channel under test (CUT) for data collection. After the transmitter modulation is completed, optical signals are coupled to the SMF for transmission by a wavelength division multiplexer. There is a pre-amplifier before entering the fiber, and an EDFA in each fiber span.

To demonstrate the generalizability of transfer learning, the signal transmission parameters such as the symbol rates, the states of polarization and the OSNR ranges are selected differently in the source domain and the target domain. The signals transmitting in the single-channel system adopted in this paper are in a low transmission symbol rate with single polarization (SP). In the single-channel system, the symbol rates of four modulation formats are unified as 10 Gbaud and the OSNR values of SP-8QAM and SP-16QAM gradually increase from 13 dB to 22 dB at 1 dB step, while the OSNR values of SP-32QAM and SP-64QAM gradually increase from 16 dB to 25 dB also at 1 dB step. To further improve the transmission system capacity, we use dual polarization (DP) signals in the three-channel WDM system. We set the symbol rates of DP-8QAM\DP-16QAM and DP-32QAM\DP-64QAM as 35 Gbaud and 25 Gbaud, respectively. The OSNR values of 8QAM and 16QAM gradually increase from 18 dB to 27 dB at 1 dB step, while the OSNR values of 32QAM and 64QAM gradually increase from 23 dB to 32 dB also at 1 dB step.

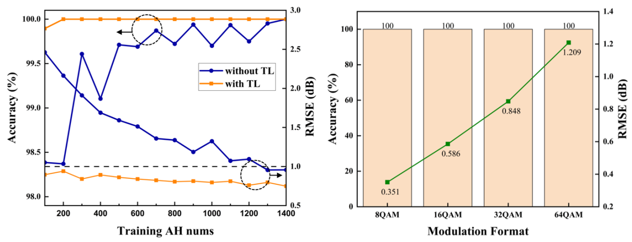
After compensation by modulation format independent algorithms such as chromatic dispersion (CD) compensation, timing recovery and constant modulus algorithm (CMA) equalization, the receiver collects 50 AHs for each OSNR value in a specific modulation format. There are a total of 2000 (=4*10*50) AHs in the dataset for each certain system setup, and these data will be fed into the proposed TL-CDNN model for off-line DSP processing. Once the modulation format has been determined successfully, the modulation format dependent algorithms can be optimized.

The biggest difference between the single-channel system and the three-channel WDM system is that the channels in the WDM system do not exist independently, but will be interfered by other channels, especially adjacent channels. In order to more comprehensively verify the applicability of the proposed method to WDM systems, we set up three scenarios to collect data (a) the laser emission power of the three channels is same; (b) the transmitting power of a channel adjacent to the CUT is 6 dBm higher and the scenario is denoted as delta_p_6; (c) the transmitting power of the CUT is 3 dBm lower than the other two adjacent channels and the scenario is denoted as CUT_-3.

B. *Results and discussion*

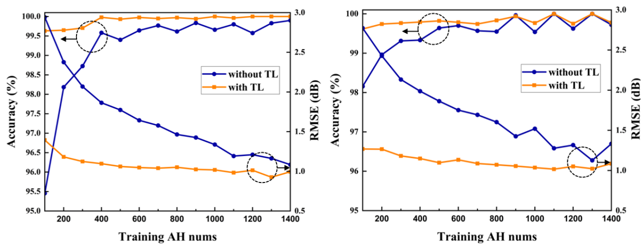
To verify the impact of transfer learning, we compare the requirements in the training samples of CDNN models with TL and without TL. When the laser emission power of the three channels is consistent, from a dataset with a total of 2000 samples in the target domain, we took 5% to 70% AHs for training at 5% growth step. As shown in Fig. 3(a), after 100 epochs, for the MFI task, both methods can achieve 100% recognition rate eventually. To converge to 100%, only 200 AHs are needed with the method using TL and 1400 AHs are required for the method without TL, respectively. For the OSNR estimation task, the error of the method with TL is always much smaller than that of the method without TL. And only 100 AHs are enough for the method with TL to reduce the root mean square error (RMSE) to lower than 1 dB, while the method without TL needs 1300 AHs to ensure this value. And Fig. 3(b) shows the accuracy and RMSE versus for four modulation formats when the amount of training AHs is sufficient. This is enough to demonstrate that the introduction of TL can significantly reduce the cost while guaranteeing the OPM performance, and only a small amount of data in the target domain to complete the OPM work.

Fig. 4 shows how the performance of CDNN models with TL and without TL varies with the training sample size when multiple channels in the WDM system transmit unequal power. Although the attenuation of the CUT power and the increase of an adjacent channel power can lead to deterioration of transmission signal quality, it is obvious in Fig. 4 that CDNN with TL always outperform the CDNN without TL. The results show that the proposed method can still perform well in MFI and OSNR monitoring when the channel is disturbed and the quality is poor.



(a) (b)

Fig. 3. (a) Accuracy and RMSE vs. the number of training AH of CDNN with TL and without TL; (b) accuracy and RMSE vs. four modulation formats with TL.



(a) (b)

Fig. 4. Accuracy and RMSE vs. the number of training AH of CDNN with TL and without TL in the case of (a) delta_p_6 and (b) CUT_-3.

IV. Conclusions

In this paper, we have proposed a cascaded deep neural network with transfer learning for the OPM tasks in a WDM system. The results for 8QAM, 16QAM, 32QAM and 64QAM signals show that using the method with TL, the proposed scheme can get a pre-trained model from a simple single-channel system to simultaneously complete the MFI and OSNR estimation task in a three-channel WDM system. By introducing TL, MFI accuracy can reach 100% and RMSE value can be ensured lower than 1 dB requiring only a small amount of training AHs collected from WDM systems. The robustness of the proposed method is also proved when the channel signal quality deteriorates due to the different signal power in different channels. TL-CDNN proposed in this paper provides further research for the combination of OPM task and deep learning based on WDM systems.

Acknowledgment

This work is supported by Innovation Program for Quantum Science and Technology 2021ZD0303200; National Natural Science Foundation of China (Grant NO. 12234014，12274132, 11904227, 12204304, 91536114, 11874152, and 11654005); Shanghai Municipal Science and Technology Major Project (Grant NO. 2019SHZDZX01); National Key Research and Development Program of China (Grant No. 2016YFA0302001); Shanghai Municipal Science and Technology Commission (Grant No. 22DZ2229004).

References

[1] L. I. Zhang, X. Zhou, J. Du, and P. Tian, "Fast self-learning modulation recognition method for smart underwater optical communication systems," *Opt. Express*, vol. 28, no. 25, pp. 38223-38240, 2020.

[2] Z. Guo et al., "Modulation format recognition with transfer learning assisted convolutional neural network using multiple Stokes sectional plane image in multi-core fibers," *Opt. Express*, vol. 30, no. 12, pp. 21990-22005, 2022.

[3] Y. Fan et al., "Experimental Validation of CNN vs. FFNN for Time- and Energy-Efficient EVM Estimation in Coherent Optical Systems," *Journal of optical communications and networking*, vol. 13, no. 10, p. E63, 2021.

[4] X. Fan, Z. Chen, R. Hao, F. Ren, and J. Wang, "Improving the adaptability of the optical performance monitor by transfer learning," *Applied optics (2004)*, vol. 60, no. 16, pp. 4827-4834, 2021.

[5] Y. Cheng, W. Zhang, S. Fu, M. Tang, and D. Liu, "Transfer learning simplified multi-task deep neural network for PDM-64QAM optical performance monitoring," *Opt. Express*, vol. 28, no. 5, pp. 7607-7617, 2020.

[6] H. Luo, Z. Huang, X. Wu, and C. Yu, "Cost-Effective Multi-Parameter Optical Performance Monitoring Using Multi-Task Deep Learning With Adaptive ADTP and AAH," *Journal of lightwave technology*, vol. 39, no. 6, pp. 1733-1741, 2021.

[7] J. Zhang et al., "Joint Modulation Format Identification and OSNR Monitoring Using Cascaded Neural Network With Transfer Learning," *IEEE photonics journal*, vol. 13, no. 1, pp. 1-10, 2021.

[8] H. Zhang, D. Zhang, and Y. L. Xue, "Constellation Diagram Analyzer Based on Few Shot Learning," in *Asia Communications and Photonics Conference*, 2021.

[9] Z. Wang, A. Yang, P. Guo, and P. He, "OSNR and nonlinear noise power estimation for optical fiber communication systems using LSTM based deep learning technique," *Opt. Express*, vol. 26, no. 16, pp. 21346-21357, 2018.

[10] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of big data*, vol. 3, no. 1, 2016.

[11] X. Zhu et al., "Transfer learning assisted convolutional neural networks for modulation format recognition in few-mode fibers," *Optics express*, vol. 29, no. 22, pp. 36953-36963, 2021.

[12] W. Mo et al., "ANN-based transfer learning for QoT prediction in real-time mixed line-rate systems," in *Optical Fiber Communication Conference*, 2018: Optica Publishing Group, p. W4F. 3.

[13] P. Zhou and J. Austin, "Learning criteria for training neural network classifiers," *Neural computing & applications*, vol. 7, pp. 334-342, 1998.

[14] D. Wallach and B. Goffinet, "Mean squared error of prediction as a criterion for evaluating and comparing system models," *Ecological modelling*, vol. 44, no. 3-4, pp. 299-306, 1989.