



# Intelligent optical performance monitor using multi-task learning based artificial neural network

ZHIQUAN WAN, ZHENMING YU,\* LIANG SHU, YILUN ZHAO, HAOJIE ZHANG, AND KUN XU

State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China

\*yuzhenming@bupt.edu.cn

**Abstract:** An intelligent optical performance monitor using multi-task learning based artificial neural network (MTL-ANN) is designed for simultaneous OSNR monitoring and modulation format identification (MFI). Signals' amplitude histograms (AHs) after constant module algorithm are selected as the input features for MTL-ANN. The results obtained from simulation and experiment of NRZ-OOK, PAM4 and PAM8 signals demonstrate that MTL-ANN could achieve OSNR monitoring and MFI simultaneously with higher accuracy and stability compared with single-task learning based ANNs (STL-ANNs). The results show an MFI accuracy of 100% for the three modulation formats under consideration. Furthermore, OSNR monitoring with mean-square error (MSE) of 0.12 dB and accuracy of 100% is achieved while regarding it as regression problem and classification problem, respectively. In this intelligent optical performance monitor, only a single MTL-ANN is deployed, which enables reduced-complexity optical performance monitor (OPM) devices for multi-parameters estimation in future heterogeneous optical network.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

With the explosive growth of traffic induced by internet of things (IOTs), artificial intelligence-intensive services and fifth-generation (5G) services, the capacity of optical communication system is increasing dramatically. To meet the large number of traffic demands, advanced optical modulation formats with high spectral efficiencies were widely studied [1]. Furthermore, reconfigurable optical add-drop multiplexers (ROADMs) together with flexible transceivers and software defined network (SDN) controllers enable elastic and reconfigurable optical network to meet a better utilization of available transmission capacity [2]. In such a network, it is essential to monitor various network performance parameters to optimize resources utilization and allocate adequate system margin [3]. Among various parameters of optical performance monitoring (OPM), optical signal-to-noise ratio (OSNR) is one of the most importance since it is directly related to bit-error ratio (BER). It also plays a vital role for fault diagnosis as well as links health detection. Existing OSNR monitoring techniques include error vector magnitude (EVM) [4], asynchronous delay-tap plots (ADTPs) [5–7], Stokes parameters [8] and intermediate frequency analysis [9].

Besides OSNR monitoring, modulation format identification (MFI) has drawn a great interest with the development of flexible transceivers and elastic optical networks (EONs) [10]. The variation of transmission parameters in EONs imposes the adaption of digital signal processing (DSP) algorithm [11]. For example, some essential equalization algorithms used in the receiver must be suitable for the modulation format of the received signals [12]. By implementing MFI, reconfigurable DSP flow can be deployed when different advanced modulation formats signals are transmitted, like PAM- $N$  transmission in intensity modulation with direct detection (IM/DD) system [13]. In this way, modulation formats can be adjusted

adaptively based on channel condition and traffic demands, which makes the optical network software-programmable [14]. Amplitude histograms (AHs) [15], ADTPs [6,7] and digital frequency-offset [16] can be selected as metrics for MFI.

Recently, with the leap-forward development of computing resource, deep learning (DL) has achieved great success in the area of computer vision (CV), natural language processing (NLP), recommend system [17,18] etc. As a current research hotpot, several DL architectures are applied in OPM to improve the monitoring accuracy. Convolutional neural networks (CNNs), which achieve tremendous successes in CV area, were adopted for OSNR monitoring and MFI in [19,20]. Eye diagrams and constellations were selected as processing objectives for IM/DD system and coherent detection system, respectively. In [21], CNN combined with asynchronously sampled data successfully achieved OSNR estimation. In [14], MFI was achieved by using CNN combined with parameters in 2D Stokes planes. In [22], long short-term memory (LSTM) network was deployed for OSNR monitoring. In [15], AHs of direct detected signals was selected as input features of artificial neural network (ANN) for MFI. In [11], the authors extended the work in [15]. AHs after constant module algorithm (CMA) were used to monitor OSNR and identify modulation format simultaneously with four ANNs. One was employed for MFI purpose and other three were employed to estimate OSNR for individual modulation formats (QPSK, 16QAM, 64QAM).

In this paper, we propose a novel intelligent optical performance monitor using multi-task learning (MTL) based ANN. MTL is an approach to improve model generalization by using the information contained in the related tasks as an inductive bias [23]. It is widely used in CV area to improve pattern identification accuracy [24]. Apart from ANN, this MTL based method can also be applied in other DL architectures like CNN and LSTM. By employing this MTL-ANN in conjunction with signals' AHs after CMA, simultaneous OSNR monitoring and MFI are achieved. In our work, we investigate the performance of OSNR monitoring while regarding it as regression problem and classification problem, respectively. A simulation system is conducted to generate three widely-used modulation formats in IM/DD system: non-return-to-zero on-off keying (NRZ-OOK), 4-ary pulse amplitude modulation (PAM4) and 8-ary pulse amplitude modulation (PAM8) signals. The results show that our proposed monitor is robust with the appearance of chromatic dispersion (CD). Moreover, an experimental system is also setup to demonstrate the performance of our proposed intelligent optical performance monitor. The results obtained from simulation and experiment show that MTL-ANN can simultaneously realize OSNR monitoring and MFI with higher accuracy and stability, compared with single-task learning based ANNs (STL-ANNs). The results show an MFI accuracy of 100% for the three modulation formats under consideration. Besides, OSNR monitoring with mean-square error (MSE) under 0.12 dB and accuracy of 100% is achieved, respectively. Finally, the computation time for testing this intelligent optical performance monitor is investigated. In this intelligent optical performance monitor, only a single MTL-ANN is deployed which enables reduced-complexity OPM devices for multi-parameters estimation.

## 2. Operating principle

In this paper, CMA equalization is adopted to compensate linear impairments in our system since it works on modulation format unassisted mode [25]. AHs with 100 bins for three signal types considered in this work are shown in Fig. 1 for four different OSNRs. It is clear from Fig. 1 that the AHs depend on modulation format as well as OSNR, thus AHs can be exploited for simultaneous OSNR monitoring and MFI. Occurrences at each bin of AHs are selected as the input features for ANN as shown in Fig. 2. The neuron numbers in input layer are equal to the bin numbers. In Fig. 2, each circle represents a neuron and it can be modeled as a logistic unit. The output of  $r$ -th neuron can be expressed as:

$$A_r(x) = f\left(\sum_i w_{ri}x_i\right) \quad (1)$$

Where  $x_i$  is the  $i$ -th input to the neuron,  $w_{ri}$  is the corresponding weight for the  $r$ -th neuron and  $f(x)$  is the activation function of the neuron. ANN models the nonlinear problem with the help of multi-neurons architecture and nonlinear activation function in each neuron. In this paper, tangent sigmoid (Tanh-sigmoid) function is selected as activation function for neurons in hidden layers. For output layer 1, which focus on MFI, softmax function is selected as activation function. The output of softmax function is a categorical probability distribution, which reveals the probability that any of the classes are true. For example, a three-output ANN with a column vector  $h_w(x) = [0.8, 0.15, 0.05]^T$  means that the output has 80 percent probability belong to the first class. For output layer 2, which focus on OSNR monitoring, we investigate two schemes, one is regarding it as a regression problem and another is regarding it as a classification problem. Linear function and softmax function are selected as activation function, respectively. Tanh-sigmoid function and softmax function are expressed as Eq. (2) and Eq. (3), respectively.

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2)$$

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_i e^{x_i}} \quad (3)$$

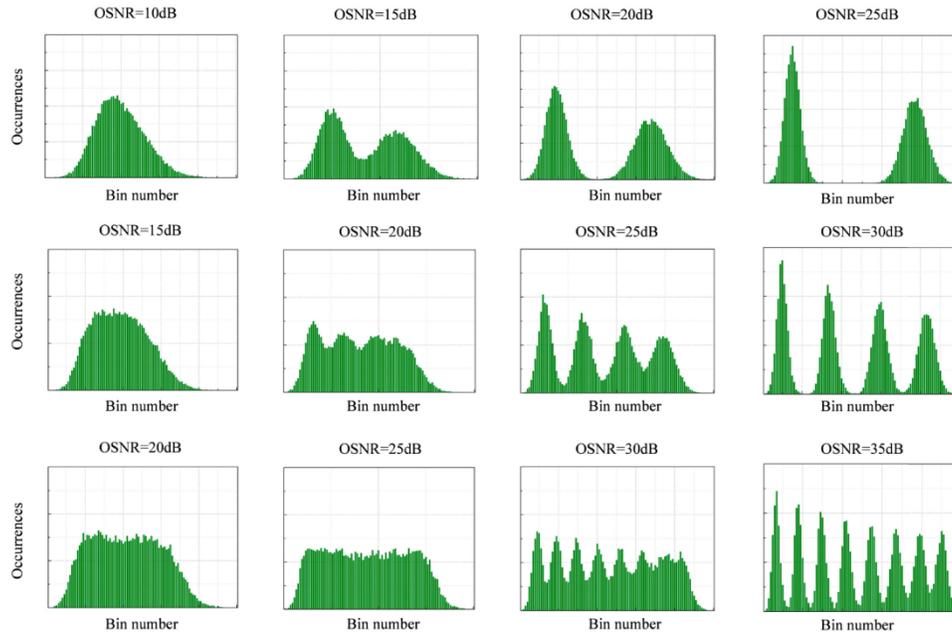


Fig. 1. AHs with 100 bins for different OSNRs for NRZ-OOK (first row), PAM4 (second row) and PAM8 (third row) signals after CMA equalization.

For an inductive learner (neural network is an inductive learner) trained by finite samples, inductive bias can cause it to prefer some hypotheses than other hypotheses. MTL uses the training samples of related tasks as an inductive bias, so the inductive learner is biased to prefer hypotheses that have utility across multi-tasks [23]. Different with STL-ANN, MTL-ANN has multi-output layers for multi-tasks as shown in Fig. 2. Since multi-tasks share a

common hidden layer, the compute resources are reduced and the commonality of different tasks can be discovered at same time. With the help of specific hidden layers for different tasks, the characteristics of different tasks are learned. The number of hidden layers and neurons in each layer can be designed for specific task. In this paper, we define the total loss function of MTL-ANN as the weighted sum of different tasks' loss function as Eq. (4) shows. In this way, effect of different tasks to the network structure is considered.

$$J(w) = \sum_{t=1}^T \left\{ w_t \sum_{k=1}^{K_t} \left| [y_t(n)]_k - [h_t(x(n))]_k \right|^2 \right\} + \lambda \sum_{i=1}^m \vartheta_i^2 \quad (4)$$

Where  $T$  represents the number of tasks,  $K_t$  represents the number of classes in  $t$ -th task,  $[h_t(x(n))]_k$  is the output of MTL-ANN belongs to class  $k$  of  $t$ -th task and  $[y_t(n)]_k$  is the reference value of  $x(n)$  which belongs to the class  $k$  of  $t$ -th task. To consider the effect of different tasks to the network,  $w_t$  is set as the loss weight of  $t$ -th task. In this paper, MFI is selected as task 1, OSNR monitoring is selected as task 2, so  $T = 2$ ,  $K_1 = 3$ ,  $K_2 = 1$  (regard as regression problem) or  $K_2 = 26$  (regard as classification problem with 26 OSNR values). To avoid overfitting problem may occur in the investigation, L2 regularization is adopted.  $\lambda$  is the coefficient of L2 regularization,  $\vartheta_i$  is the parameters of ANN (i.e. weights and bias) and  $m$  represents the number of parameters. After network structure selected, Adam algorithm is adopted to adjust weights in MTL-ANN. Compared to traditional stochastic gradient descent algorithm, Adam algorithm is computationally efficient and requires little memory [26].

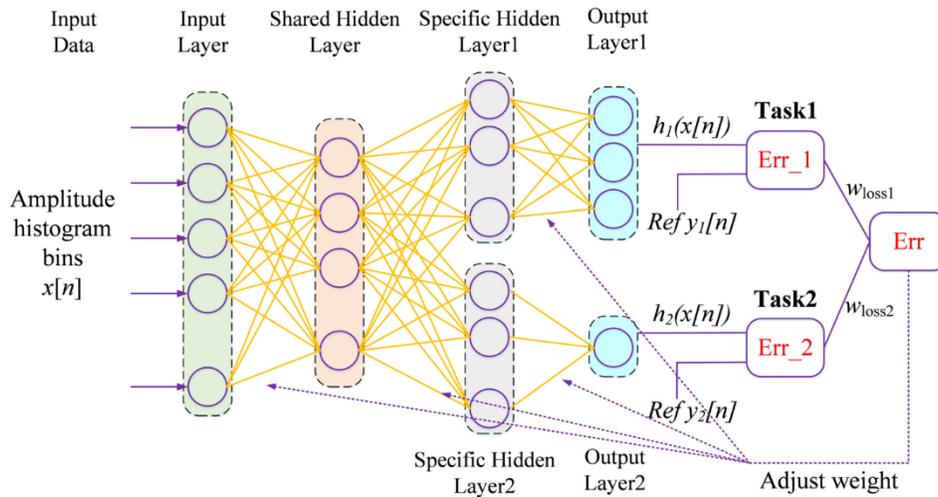


Fig. 2. Schematic structure of MTL-ANN.

### 3. MTL-ANN detection system and results

The simulation system setup based on VPI transmission Maker 9.1 is shown in Fig. 3. Three widely-used optical signals in IM/DD system (NRZ-OOK, PAM4, PAM8) with a pattern length of  $2^{13}-1$  symbols at 28 GBaud are generated at transmitter side. After modulated to the Mach-Zehnder modulator (MZM), a chromatic dispersion (CD) emulator is used to simulate the real optical signals. The CD is ranging from  $-100$  ps/nm to  $100$  ps/nm at a step of  $10$  ps/nm. After that, an OSNR setting module is used to add amplifier spontaneous emission noise (ASE) to reach the specified OSNR. The bandwidth of MZM and receiver are  $32$  GHz without considering bandwidth limitation. After detected by a PIN photodetector, the electrical signal is synchronous sampled by a  $160$  GSa/s analog to digital converter (ADC).

At the beginning of the offline digital signal processing (DSP) flow, we first remove the direct-current (DC) offset. After that, the data stream is resampled to two samples per symbol to enable the proposed equalization algorithm. After CMA-based equalization, linear transmission impairments is compensated. The AHs with specific bins are generated from the equalized samples. The occurrences at each bin are selected as input features for the following ANN. To investigate the efficiency of MTL-ANN, STL-ANN with identical network structure is employed as a comparison. For example, when dealing with task 1, specific hidden layer 2 and output layer 2 in Fig. 2 are discarded. After obtaining OSNR information, fault diagnosis and links health detection can be deployed. On the other hand, MFI information can be used for modulation format assisted equalization algorithm, like decision-directed least-mean-square (DD-LMS) algorithm based volterra nonlinear equalizer (VNLE) [12], to improve system BER performance. In this paper, Keras library combined with Tensorflow backend are selected as the model of ANN [27].

Based on the above system, we collect 10 AHs for each OSNR (10 dB-25 dB for NRZ-OOK, 15 dB-30 dB for PAM4, 20 dB-35 dB for PAM8) at specific CD value of each modulation format. The entire data set comprises 10080 (16 x 21 x 10 x 3) AHs in total. The AHs in this data set are divided into training and testing sets by random selecting 90% (i.e. 9072) and 10% (i.e. 1008) of all AHs, respectively. In our work, apart from L2 regularization, cross-validation set is also deployed to avoid overfitting problem. 10% of the data (i.e. 907) in the training set are selected as cross-validation set.

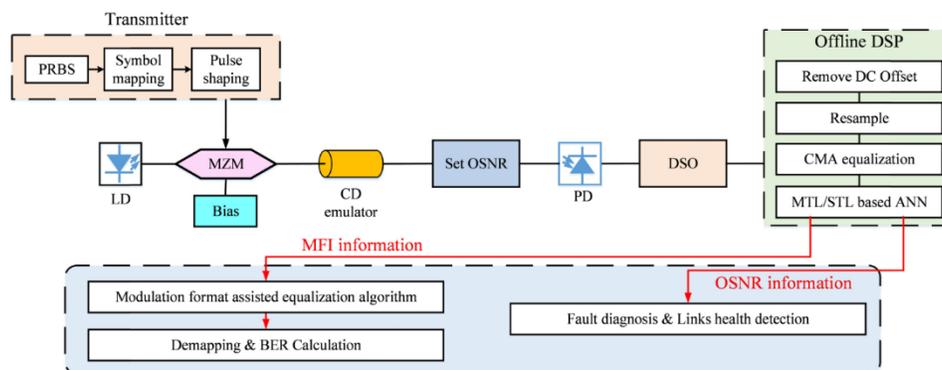


Fig. 3. Simulation setup for simultaneous OSNR monitoring and MFI.

At first, OSNR monitoring is regarded as a regression problem. We first optimize the neuron number in shared hidden layer. The bin numbers of AHs are set as 100, which means the neuron number in input layer is 100. The neurons in specific hidden layer 1, 2 are designed a half of the neurons in shared hidden layer respectively. The MFI accuracy and OSNR estimated mean square error (MSE) versus neurons in shared hidden layer for STL-ANN and MTL-ANN are shown in Fig. 4. The ratio of OSNR loss weight to MFI loss weight is set as 1 while investigating the hyperparameters (i.e., number of hidden layers, neuron numbers in each layer) of ANN. Since the performance of ANN is affected by the random initialization of ANN weights, we evaluate the performance by taking average value, maximum value and minimum value from eight random initialization. As shown in Fig. 4(a), the MFI accuracy of three modulation formats under consideration is 100% when neuron number is larger than 10 for MTL-ANN. As to STL-ANN, the average value of MFI accuracy can't reach 100% no matter how many neurons in shared hidden layer used. Besides that, with the increasement of neuron number, MFI accuracy is more sensitive to the random initialization of ANN weights. Figure 4(b) shows the OSNR estimated MSE versus neurons in shared hidden layer. With a small neuron number, OSNR estimation problem can't be model precisely. However, when neuron number is too large, more data is needed to

train the network. In our simulation, the optimal neuron numbers in shared hidden layer for STL-ANN and MTL-ANN are both about 40.

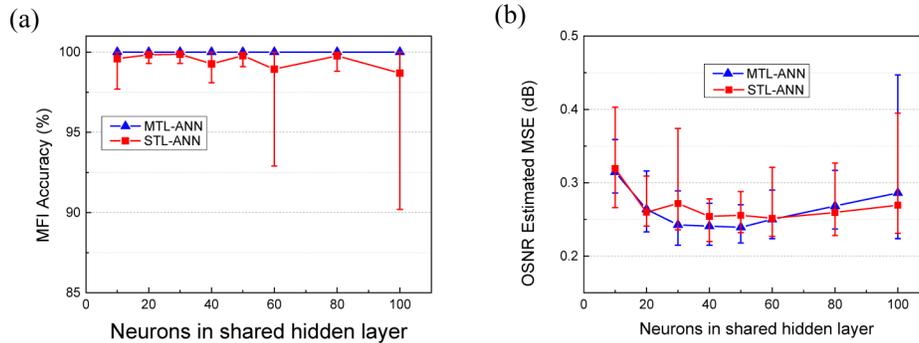


Fig. 4. (a) MFI accuracy and (b) OSNR estimated MSE versus hidden neurons in shared hidden layer for MTL-ANN and STL-ANN (Average, maximum and minimum value from eight random initialization).

After optimizing neuron number in shared hidden layer, we investigate the optimal loss weight ratio. Figure 5 shows OSNR estimated MSE versus ratio of OSNR loss weight to MFI loss weight at optimal network hyperparameters. Loss weights of different tasks are defined in Fig. 2. As can be seen from Fig. 5, the OSNR estimated MSE decrease monotonically when the ratio is less than 100. Which means the OSNR monitoring task plays a more important role than MFI task in our designed MTL-ANN. The red dotted line represents the optimal average OSNR estimated MSE of STL-ANN. When the loss weight ratio of OSNR to MFI is larger than 1, the OSNR monitoring performance of MTL-ANN is better than STL-ANN. The optimal loss weight ratio of OSNR to MFI is 100 and the OSNR estimated MSE is 0.12 dB at this situation. Figure 6 shows the estimated OSNRs versus true OSNRs of MTL-ANN at optimal shared hidden layer neurons and loss weight ratio. As can be seen from the Fig. 6, our designed MTL-ANN works well in the OSNR range (10 dB-35 dB) under consideration.

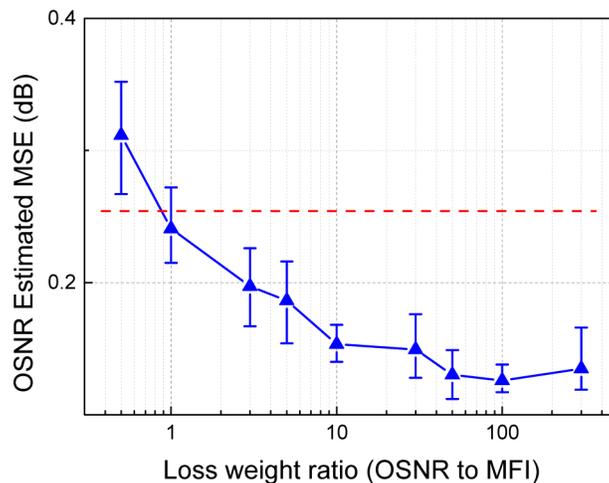


Fig. 5. OSNR estimated MSE versus ratio of OSNR loss weight to MFI loss weight (Average, maximum and minimum value from eight random initialization).

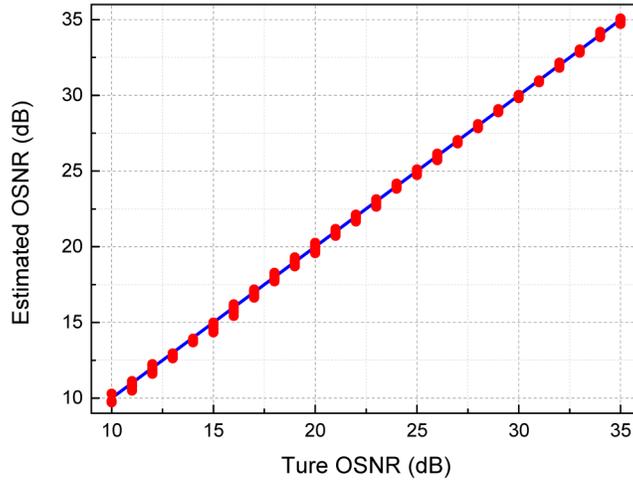


Fig. 6. True OSNRs versus estimated OSNRs of MTL-ANN.

Besides regarding OSNR monitoring as a regression problem, we also investigate the performance of regarding OSNR monitoring as a classification problem, since OSNR monitoring with 1 dB resolution is enough for real situation. In that case, 26 OSNR values (from 10 dB to 35 dB) requiring a one-hot vector with 26 elements to represent the corresponding OSNR values. Figure 7 shows MFI accuracy and OSNR accuracy versus neurons in shared hidden layer for STL-ANN and MTL-ANN, respectively. As shown in Fig. 7(a), the MFI accuracy of three modulation formats under consideration is also 100% while regarding OSNR monitoring as a classification problem for MTL-ANN. In Fig. 7(b), we find OSNR accuracy can reach 100% with small vibration when neuron number is 60 for STL-ANN and MTL-ANN with OSNR to MFI loss weight ratio is 1. By adopting MTL-ANN with OSNR to MFI loss weight ratio is 100, the OSNR accuracy can reach 100% with no vibration when neuron number is larger than 20. To sum up, we find MTL-ANN could achieve OSNR monitoring and MFI simultaneously with higher accuracy and stability compared with STL-ANNs, no matter OSNR monitoring is regarded as a regression problem or classification problem.

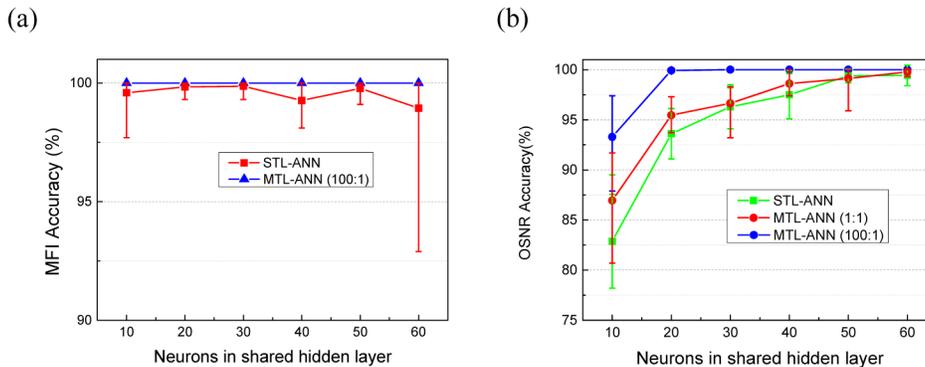


Fig. 7. (a) MFI accuracy versus hidden neurons in shared hidden layer (b) OSNR accuracy versus hidden neurons in shared hidden layer for STL-ANN and MTL-ANN with different loss ratio (Average, maximum and minimum value from eight random initialization).

#### 4. Experimental setup and results

To demonstrate the feasibility and effectiveness of proposed scheme, we also set up a physical experiment. The experimental setup is shown in Fig. 8. The three modulation formats under consideration with a pattern length of  $2^{13}-1$  symbols at 20 GBaud are generated by a 50 GSa/s arbitrary waveform generator (AWG, Tektronix AWG70001A). The signal with a peak-to-peak voltage of  $\sim 1V$  drives the Mach-Zehnder modulator (MZM). Different with simulation system which only consider CD effecting, the generated optical signal in experimental setup is launched to a spool of dispersion-uncompensated standard single mode fiber (SSMF) with 5-km fiber reach. The launched power is set as 1.5 dBm. After fiber transmission, an Erbium-doped fiber amplifier (EDFA) and a variable optical attenuator (VOA) are employed to load optical noise and adjust the OSNR from 14 dB to 29 dB for NRZ-OOK signal, 17 dB to 32 dB for PAM signal and 22 dB to 37 dB for PAM8 signal at step of 1 dB. At the receiver, the signal is passed through a 1-nm optical band pass filter (OBPF). The resulting OSNR is measured by an optical spectrum analyzer (OSA). After detected by a PIN photodetector, the electrical signal is sampled by a 100 GSa/s digital sampling oscilloscope (DSO, Tektronix MSO72004C). The reference clock output port of AWG and the external reference clock input port of DSO are connected to synchronize the clock. Finally, the digital signals are processed by the proposed offline DSP.

Based on the above system, we collect 100 AHs for each OSNR value of each modulation format. The entire data set comprises 4800 ( $100 \times 16 \times 3$ ) AHs in total. The AHs in this data set are divided into training and testing sets by random selecting 90% (i.e. 4320) and 10% (i.e. 480) of all AHs, respectively. In the training sets, 10% of the data (i.e. 432) are selected as cross-validation set.

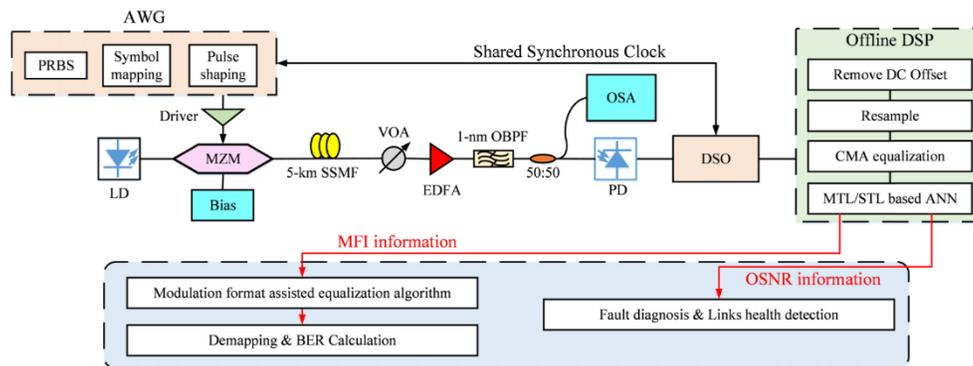


Fig. 8. Experimental setup for simultaneous OSNR monitoring and MFI.

At first, OSNR monitoring is also regarded as a regression problem. Compared to simulation setup, more impairments are existed in experimental setup. In that way more bins are needed to describe AH, so the neuron number in input layer is set as 150. The MFI accuracy and OSNR estimated MSE versus neurons in shared hidden layer for STL-ANN and MTL-ANN are shown in Fig. 9. Based on the simulation results, the loss weight ratio of OSNR to MFI is set as 100. As shown in Fig. 9(a), MFI accuracy of three modulation formats under consideration is 100% when neuron number is larger than 20 for MTL-ANN. As to STL-ANN, MFI accuracy suffers larger vibration compared to the results in simulation system. Figure 9(b) shows the OSNR estimated MSE versus neurons in shared hidden layer. The optimal neuron numbers in shared hidden layer for STL-ANN and MTL-ANN are 60 and 50, respectively. The optimal OSNR estimated MSE is 0.4 dB for STL-ANN, larger than the simulation results. However, the OSNR estimated MSE for MTL-ANN with 100 loss

ratio is less than 0.4 dB and the optimal estimated MSE is 0.11 dB, same like simulation results.

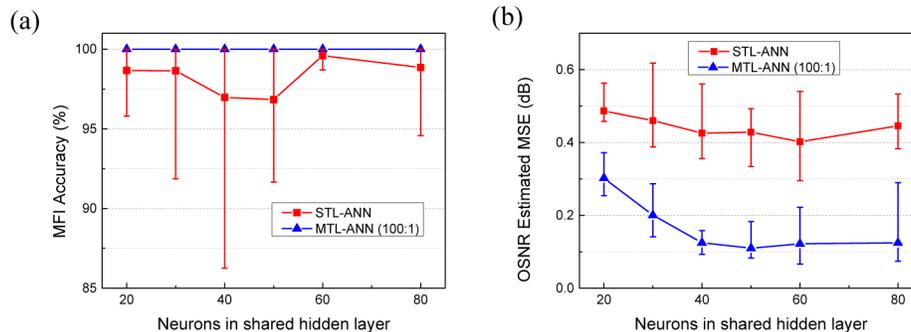


Fig. 9. Experimental (a) MFI accuracy and (b) OSNR estimated MSE versus hidden neurons in shared hidden layer for STL-ANN and MTL-ANN with 100 loss ratio (Average, maximum and minimum value from eight random initialization).

After regarding OSNR monitoring as a regression problem for the experimental data, we also investigate the performance of regarding OSNR monitoring as a classification problem. A one-hot vector with 24 elements (from 14 dB to 37 dB) is selected to represent the corresponding OSNR values. Figure 10 shows MFI accuracy and OSNR accuracy versus neurons in shared hidden layer for STL-ANN and MTL-ANN, respectively. Same like simulation results, the MFI accuracy of three modulation formats under consideration is 100% while regarding OSNR monitoring as a classification problem for MTL-ANN. As shown in Fig. 10(b), by adopting MTL-ANN with OSNR to MFI loss weight ratio is 100, OSNR accuracy can reach 100% without vibration when neuron number is larger than 60. For STL-ANN, OSNR accuracy can't reach 100% without vibration no matter how many neurons selected.

Finally, we investigate the computation time per sample for testing MTL-ANN and STL-ANNs while regarding OSNR monitoring as regression and classification problem, respectively. Experimental data with optimal ANN hyperparameters is selected to test the computation time. The ANNs are running in the laptop with Intel Core i7-7500U CPU @ 2.70GHz and 2.90GHz. In Table 1, the network structure for STL-ANNs represents the neuron number selected for MFI and OSNR monitoring, respectively. As can be seen from Table 1, the computation time is almost proportionate to the scale to ANN. The computation time is less than 0.02ms in the proposed intelligent optical performance monitor. Besides that, the computation time reduced about 55% and 41% while considering OSNR monitoring as regression and classification, respectively.

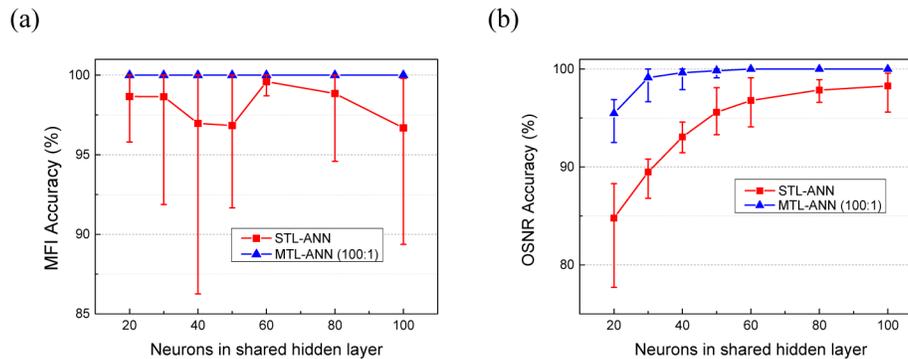


Fig. 10. Experimental (a) MFI accuracy versus hidden neurons in shared hidden layer (b) OSNR accuracy versus hidden neurons in shared hidden layer for STL-ANN and MTL-ANN with 100 loss ratio (Average, maximum and minimum value from eight random initialization).

Table 1. Computation time per sample for testing MTL-ANN and STL-ANNs

Network type	Network structure (shared layer + specific layer)	Computation time (ms)
MTL-ANN for regression problem	150 + 50	0.01662
STL-ANNs for regression problem	150 + 20, 150 + 60	0.03740
MTL-ANN for classification problem	150 + 50	0.02078
STL-ANNs for classification problem	150 + 20, 150 + 60	0.03532

## 5. Conclusion

In this paper, we have proposed an intelligent optical performance monitor using MTL-ANN to simultaneously monitor OSNR and identify modulation format with signals' AHs. OSNR monitoring is regarded as regression problem and classification problem in this monitor, respectively. The simulated results and experimental results show that MFI accuracy of three widely-used modulation formats in IM/DD system under consideration is 100% in estimated OSNR range for both case. Besides, OSNR estimation with MSE under 0.12 dB and accuracy of 100% are achieved while regarding it as a regression problem and classification problem, respectively. By choosing the weighted sum of different tasks' loss function as total loss function, the effect of different tasks to the network is considered. In our work, OSNR monitoring task plays a more important role than MFI task for designed MTL-ANN. Compared to STL-ANNs, MTL-ANN improves the performance of MFI and OSNR monitoring. Besides that, the performance vibration induced by ANN random initialization is overcome. In this intelligent optical performance monitor, only a single MTL-ANN is deployed to estimate multi-parameters. Compared to STL-ANNs, the computation time reduced more than 40%. With more parameters to estimate, the computation time can be further reduced, which is attractive for real-time multi-parameters estimation in future heterogeneous optical network.

## Funding

NSFC Program (No. 61431003, 61601049, 61625104, 61821001); BUPT Excellent Ph.D. Students Foundation (CX2018115); Beijing Municipal Science & Technology Commission (Grant No. Z181100008918011); Fund of State Key Laboratory of Information Photonics and Optical Communications, BUPT (No. IPOC2017ZT08).

## References

1. A. P. T. Lau, Y. Gao, Q. Sui, D. Wang, Q. Zhuge, M. H. Morsy-Osman, M. Chagnon, X. Xu, C. Lu, and D. V. Plant, "Advanced DSP techniques enabling high spectral efficiency and flexible transmissions: toward elastic optical networks," *IEEE Signal Process. Mag.* **31**(2), 82–92 (2014).

2. S. Oda, M. Bouda, Y. Ge, S. Yoshida, T. Tanimura, Y. Akiyama, Y. Hirose, Z. Tao, T. Lkeuchi, and T. Hoshida, "Innovative Optical Networking by Optical Performance Monitoring and Learning Process," in *European Conference on Optical Communication (ECOC)* (2018), p. 1.
3. Z. Dong, F. N. Khan, Q. Sui, K. Zhong, C. Lu, and A. P. T. Lau, "Optical performance monitoring: a review of current and future technologies," *J. Lightwave Technol.* **34**(2), 525–543 (2016).
4. Z. Dong, A. P. T. Lau, and C. Lu, "OSNR monitoring for QPSK and 16-QAM systems in presence of fiber nonlinearities for digital coherent receivers," *Opt. Express* **20**(17), 19520–19534 (2012).
5. M. C. Tan, F. N. Khan, W. H. Al-Arashi, Y. Zhou, and A. P. T. Lau, "Simultaneous optical performance monitoring and modulation format/bit-rate identification using principal component analysis," *J. Opt. Commun. Netw.* **6**(5), 441 (2014).
6. F. N. Khan, Y. Yu, M. C. Tan, W. H. Al-Arashi, C. Yu, A. P. T. Lau, and C. Lu, "Experimental demonstration of joint OSNR monitoring and modulation format identification using asynchronous single channel sampling," *Opt. Express* **23**(23), 30337–30346 (2015).
7. X. Fan, Y. Xie, F. Ren, Y. Zhang, X. Huang, W. Chen, T. Zhangsun, and J. Wang, "Joint optical performance monitoring and modulation format/bit-rate identification by CNN-based multi-task learning," *IEEE Photonics J.* **10**(5), 1–12 (2018).
8. L. Lundberg, H. Sunnerud, and P. Johannisson, "In-Band OSNR Monitoring of PM-QPSK Using the Stokes Parameters," in *Optical Fiber Communication Conference (OSA)*, (2015), p. W4D.5.
9. L. Baker-Meflah, B. Thomsen, J. Mitchell, and P. Bayvel, "Simultaneous chromatic dispersion, polarization-mode-dispersion and OSNR monitoring at 40Gbit/s," *Opt. Express* **16**(20), 15999–16004 (2008).
10. I. Tomkos, S. Azodolmolky, J. Sole-Pareta, D. Careglio, and E. Palkopoulou, "A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges," *Proc. IEEE* **102**(9), 1317–1337 (2014).
11. F. N. Khan, K. Zhong, X. Zhou, W. H. Al-Arashi, C. Yu, C. Lu, and A. P. T. Lau, "Joint OSNR monitoring and modulation format identification in digital coherent receivers using deep neural networks," *Opt. Express* **25**(15), 17767–17776 (2017).
12. Z. Wan, J. Li, L. Shu, S. Fu, Y. Fan, F. Yin, Y. Zhou, Y. Dai, and K. Xu, "64-Gb/s SSB-PAM4 transmission over 120-km dispersion-uncompensated SSMF with blind nonlinear equalization, adaptive noise-whitening postfilter and MLSDF," *J. Lightwave Technol.* **35**(23), 5193–5200 (2017).
13. F. Li, D. Zou, Q. Sui, J. Li, X. Yi, L. Li, and Z. Li, "Optical Amplifier-free 100 Gbit/s/lamda PAM-N Transmission and Reception in O-band over 40-km SMF with 10-G Class DML," in *Optical Fiber Communication Conference (OFC) 2019 (OSA)*, (2019), p. Tu2F.4.
14. W. Zhang, D. Zhu, Z. He, N. Zhang, X. Zhang, H. Zhang, and Y. Li, "Identifying modulation formats through 2D Stokes planes with deep neural networks," *Opt. Express* **26**(18), 23507–23517 (2018).
15. F. N. Khan, Y. Zhou, A. P. T. Lau, and C. Lu, "Modulation format identification in heterogeneous fiber-optic networks using artificial neural networks," *Opt. Express* **20**(11), 12422–12431 (2012).
16. S. Fu, Z. Xu, J. Lu, H. Jiang, Q. Wu, Z. Hu, M. Tang, D. Liu, and C. C.-K. Chan, "Modulation format identification enabled by the digital frequency-offset loading technique for hitless coherent transceiver," *Opt. Express* **26**(6), 7288–7296 (2018).
17. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM* **60**(6), 84–90 (2017).
18. M. M. Lopez and J. Kalita, "Deep Learning applied to NLP," arXiv:1703.03091[cs] (2017).
19. D. Wang, M. Zhang, Z. Li, J. Li, M. Fu, Y. Cui, and X. Chen, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photonics Technol. Lett.* **29**(19), 1667–1670 (2017).
20. D. Wang, M. Zhang, J. Li, Z. Li, J. Li, C. Song, and X. Chen, "Intelligent constellation diagram analyzer using convolutional neural network-based deep learning," *Opt. Express* **25**(15), 17150–17166 (2017).
21. T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, "Data-analytics-based Optical Performance Monitoring Technique for Optical Transport Networks," in *Optical Fiber Communication Conference (OSA)*, (2018), p. Tu3E.3.
22. 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* **26**(16), 21346–21357 (2018).
23. R. Caruana, "Multitask Learning," *Mach. Learn.* **28**(1), 41–75 (1997).
24. M. Long, Z. Cao, J. Wang, and P. S. Yu, "Learning Multiple Tasks with Multilinear Relationship Networks," arXiv:1506.02117 [cs] (2015).
25. J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust. Speech Signal Process.* **31**(2), 459–472 (1983).
26. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv:1412.6980 [cs] (2014).
27. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," arXiv:1603.04467 [cs] (2016).