



# Loss weight adaptive multi-task learning based optical performance monitor for multiple parameters estimation

ZHENMING YU,<sup>1</sup> ZHIQUAN WAN,<sup>1,3</sup> LIANG SHU,<sup>1</sup>  SHAOHUA HU,<sup>2</sup> YILUN ZHAO,<sup>1</sup> JING ZHANG,<sup>2</sup> AND KUN XU<sup>1,4</sup>

<sup>1</sup>State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing, 100876, China

<sup>2</sup>Key Laboratory of Optical Fiber Sensing and Communications, University of Electronic Science and Technology of China, Chengdu, 611731, China

<sup>3</sup>wanzhiquan@bupt.edu.cn

<sup>4</sup>xukun@bupt.edu.cn

**Abstract:** A loss weight adaptive multi-task learning based artificial neural network (MTL-ANN) is applied for joint optical signal-to-noise ratio (OSNR) monitoring and modulation format identification (MFI). We conduct an experiment of polarization division multiplexing (PDM) coherent optical system with 5 km standard single mode fiber (SSMF) transmission to verify this monitor. A group of modulation schemes including nine modulation adaptive M-QAM formats are selected as the transmission signals. Instead of circular constellation, signals' amplitude histograms after constant module algorithm (CMA) based polarization de-multiplexing are selected as input features for our proposed monitor. The experimental results show that the MFI accuracy reaches 100% in the estimated OSNR range. Furthermore, when treated as regression problem and classification problem, OSNR estimation with a root mean-square error (RMSE) of 0.68 dB and an accuracy of 98.7% are achieved, respectively. Unlike loss weight fixed MTL-ANN, loss weight adaptive MTL-ANN could search the optimal loss weight ratio automatically for different link configurations. Besides that, the number of estimated parameters can be easily expanded, which is attractive for multiple parameters estimation in future heterogeneous optical networks.

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

## 1. Introduction

The Cisco visual networking index (2017-2022) shows annual global traffic will reach 4.8 ZB by 2022, which is 3.7-fold of 2017. Besides, busy hour internet traffic is growing even more rapidly, which will increase by a factor of 4.8 between 2017 and 2022 [1]. With the development of digital signal processing algorithms, advanced optical modulation formats, probabilistic constellation shaping and forward error correction, the capacity of conventional single mode fiber has approaching Shannon's limit [2]. In this situation, elastic optical networks (EONs) enabled by flexible transceivers together with reconfigurable optical add-drop multiplexers (ROADMs) and software defined network (SDN) controllers, are deployed to realize a more efficient utilization of physical layer resources from the network perspective [3,4]. In EONs, data rate, modulation format, transmission power, etc. can all adjust adaptively based on channel conditions and capacity demands for different services and ends. It is essential to monitor various network parameters to optimize resources utilization and allocate adequate system margin [5]. Consequently, optical performance monitoring (OPM) is indispensable in enabling flexibility and efficiency for EONs.

Among various parameters of OPM, optical signal-to-noise ratio (OSNR) is one of the most importance due to its direct relation with signal quality after equalization and bit-error ratio (BER)

[6]. Since OSNR is critical to ensure transmission quality, it should be monitored ubiquitously across the transmission link including intermediate nodes and destination nodes. Besides OSNR monitoring, modulation format identification (MFI) has also drawn great interest with the development of flexible transceivers and EONs [7]. It could grant autonomy and flexibility to the network owing to carrier phase recovery, frequency offset compensation, decision-directed least-mean-square (DD-LMS) and multi-modulus based equalization algorithms in receivers are modulation format dependent [8–10]. With the development of digital signal processing (DSP), OPM utilizing data signals after O/E conversion has gained substantial attention. Existing features for electrical domain-based OPM including amplitude histograms (AHs) [10,11], asynchronous delay-tap plots (ADTPs) [12,13], peak-to-average-power ratio (PAPR) of signal [8], Stokes parameters [4,14], error vector magnitude [15], digital frequency-offset [16] and intermediate frequency analysis [17]. Among these metrics, ADTPs and AHs draw more concerns since they are capable of monitoring multiple impairments simultaneously [5]. In addition to OSNR and modulation formats, several parameters are also monitored in EONs such as bit rate, chromatic dispersion (CD), polarization mode dispersion (PMD), nonlinear noise power etc. [17–19].

To simultaneously estimate multiple parameters in EONs, we proposed an intelligent optical performance monitor using multi-task learning based artificial neural network (MTL-ANN) in our previous work [11]. In intensity modulation with direct detection (IM/DD) system, OSNR monitoring and MFI were achieved jointly with higher accuracy and stability compared with single-task learning based ANNs (STL-ANNs). In [11], we demonstrated that the performance of MTL-ANN strongly depended on the relative loss weight ratio of each tasks. However, the loss weight ratio of different tasks is tuned manually, which is difficult and expensive for more than three tasks. In [20], the authors proposed a principled way to adjust loss weight automatically for MTL based network in computer vision (CV) area.

In this paper, we apply this loss weight adaptive method for the MTL-ANN based optical performance monitor for real-time multiple parameters estimation in future heterogeneous optical networks. By using this monitor, OSNR estimation and MFI for the modulation format adaptive M-QAM scheme in coherent polarization division multiplexing (PDM) system are achieved simultaneously. To reduce the complexity of monitor, ANN, rather than convolutional neural network (CNN) or long short-term memory (LSTM) network, is selected as network structure. In our work, a coherent PDM experimental system over 5 km is conducted to generate nine modulation format adaptive M-QAM. Signals' AHs after polarization de-multiplexing are selected as inputs for this monitor. The experimental results show an MFI accuracy of 100% for the nine modulation formats under consideration. Besides, OSNR monitoring with root mean-square error (RMSE) of 0.68 dB and accuracy of 98.7% are achieved when treated as regression problem and classification problem, respectively. We also build a simulation system based on VPI transmission Maker 9.1 to investigate the necessity and generalization ability of the loss weight adaptive MTL-ANN. Compared with loss weight adaptive MTL-ANN, loss weight fixed MTL-ANN with optimal loss weight ratio achieves similar MFI and OSNR monitoring performance. However, the optimal loss weight ratio varies with different link configurations. Therefore, it is time-consuming and complicated to find the optimal loss weight ratio for different link configurations. By adopting the loss weight adaptive method, the optimal ratio could adjust automatically. Furthermore, the number of estimated parameters can be easily expanded, which is attractive for multiple parameters estimation in future heterogeneous optical network.

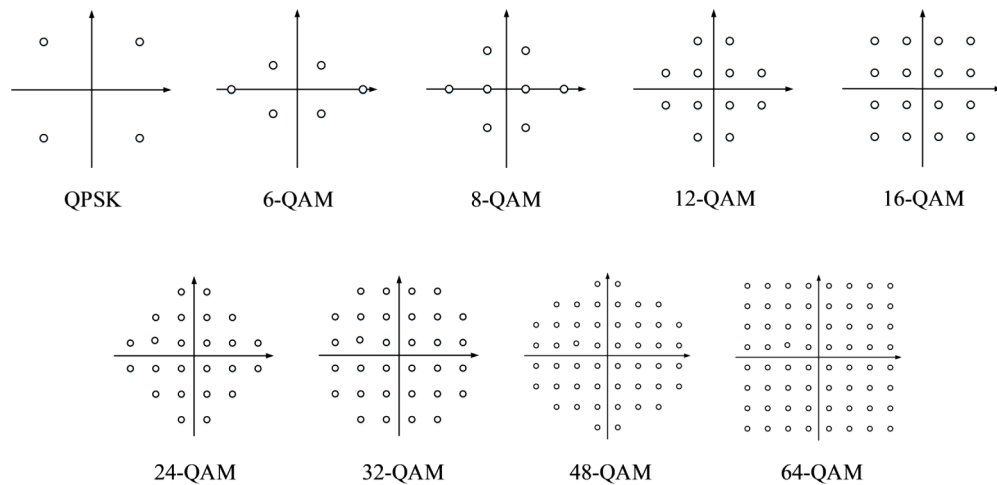
## 2. Operating principle

### 2.1. Modulation format adaptive M-QAM and AHs

Adaptive modulation increases the capacity of network by adjusting the modulation scheme according to the channel status. In coherent system, QPSK, 16-QAM and 64-QAM are most popular modulation formats. However, the gap among these three modulation scheme is too

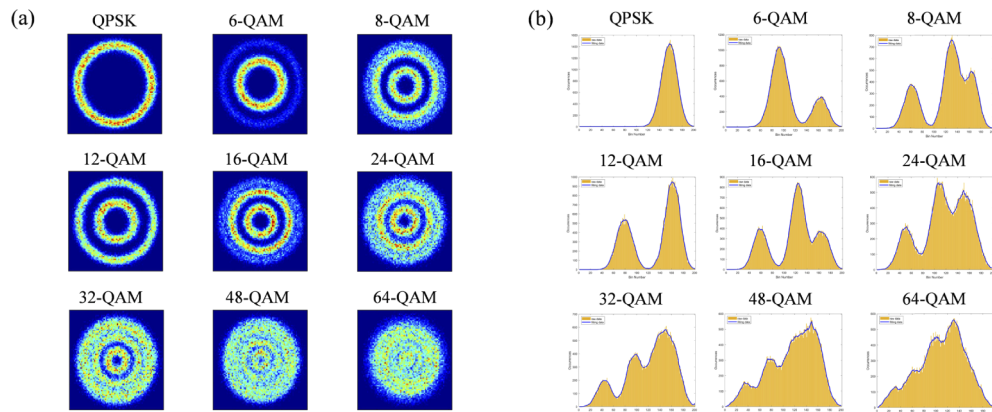
big for flexible deployment of adaptive modulation. In [21], a group of modulation schemes of M-QAM are proposed to improve capacity and flexibility of system. In schemes,  $n$ -bit QAM modulation where  $n$  ranges from 2 to 6 including half-integer indices is proposed to smoothly fill the gap. For modulation schemes with half-integer indices, two symbols are used to transfer  $2p + 1$  bits ( $p$  is integer greater than 1), meaning that a group of M-QAM with  $M = 3 \times 2^{(p-1)}$  is achieved. For example, if  $p$  equals to 4, 4.5 bits are transferred each symbol, where 24-QAM is achieved. In this way, an intermediate modulation scheme lies between 16-QAM and 32-QAM is achieved. The bit mapping scheme and configuration of M-QAM are shown in [21]. This modulation scheme is adopted in optical direct-detection OFDM system to better utilize the bandwidth [22].

Figure 1 shows the constellation points of the proposed nine modulation format adaptive M-QAM. The Euclidean distance among neighboring points is same in the designed constellation diagrams. After IQ imbalance compensation, CD compensation and modulation format independent constant module algorithm (CMA) based polarization de-multiplexing, the heat map of constellation diagrams at X-polarization with phase rotation are shown in Fig. 2(a). Different color of the diagram means the density of the constellation points at corresponding grid. Though 6-QAM signal and 12-QAM signal both have two circles, 8-QAM signal and 16-QAM signal both have three circles, the position and point number of each circle are different. Since circular constellation diagrams mainly contain amplitude information of signal, we transform the 2-D constellations into 1-D AHs with 200 bins as shown in Fig. 2(b). In this way, the data is compressed which is helpful for complexity-reduced OPM algorithm. Besides that, the difference of point number for each circle is more obvious in AHs. Instead of normalizing AHs into the whole bin area [10,11], the position information is preserved while normalizing AHs. In this way, more information in constellation diagrams is preserved, which is helpful to achieve OSNR monitoring and MFI with higher accuracy. To overcome the small fluctuation in AH induced by the incomplete compensation of channel impairments, curve fitting with the average of neighboring five amplitudes is adopted as Fig. 2(b) blue line shows.

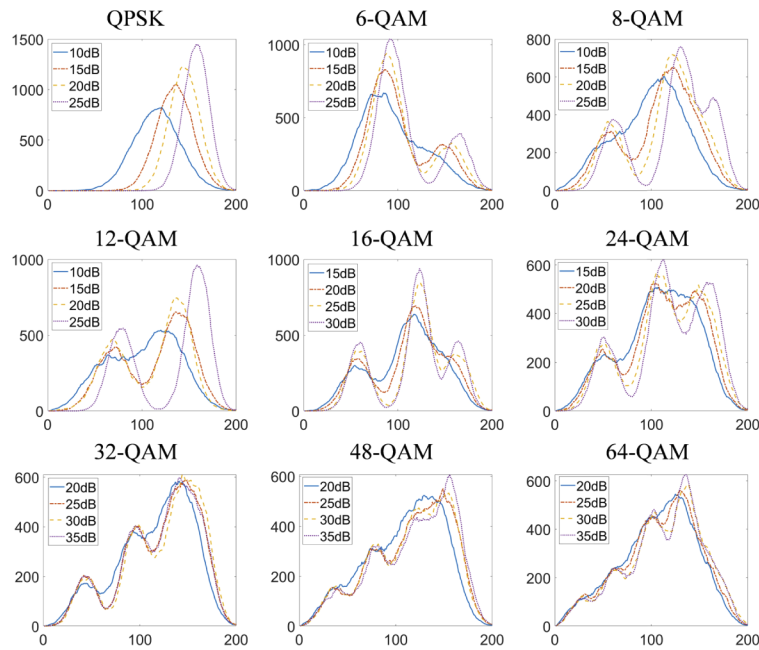


**Fig. 1.** Nine schematic constellation diagrams of modulation format adaptive M-QAM scheme.

Figure 3 shows the AHs after curve fitting for nine modulation formats at different OSNR. It is clear from Fig. 3 that AHs depend on modulation format as well as OSNR, thus AHs can be exploited for joint OSNR monitoring and MFI.



**Fig. 2.** Modulation format adaptive M-QAM schemes. (a) Heat map of constellation diagrams at X-polarization with an OSNR of 25 dB after polarization de-multiplexing, (b) Amplitude histograms (AHs) of constellation diagrams.



**Fig. 3.** AHs after curve fitting for nine modulation formats at different OSNR (10, 15, 20, 25 dB for QPSK, 6-QAM, 8-QAM, 12-QAM; 15, 20, 25, 30 dB for 16-QAM, 24-QAM; 20, 25, 30, 35 dB for 32-QAM, 48-QAM, 64-QAM).

## 2.2. Loss weight adaptive MTL-ANN

MTL can be considered as an approach of inductive knowledge transfer which improves generalization by sharing domain information among related tasks [23]. It does that by using a shared representation to learn multiple tasks, which means those learned from one task can help learn other tasks. It can also be considered as one of transfer learning. It is firstly proposed in CV area to improve pattern identification accuracy [24]. The schematic structure of MTL-ANN is shown in Fig. 4, multiple output layers are deployed for multiple tasks. Common hidden layers

and specific layers are deployed to discover the commonality and characteristics of different tasks, respectively. Number of specific layers and neurons in specific layers can be designed individually for different tasks. For each task, it could be either classification problem or regression problem. Since the performance of MTL-ANN strongly depends on the relative loss weight of each tasks [11], it is important to find an optimal loss weight. However, searching an optimal weight is prohibitively expensive and difficult with manual tuning. In this paper, we use homoscedastic uncertainty mentioned in [20] to combine multiple loss functions of each tasks.

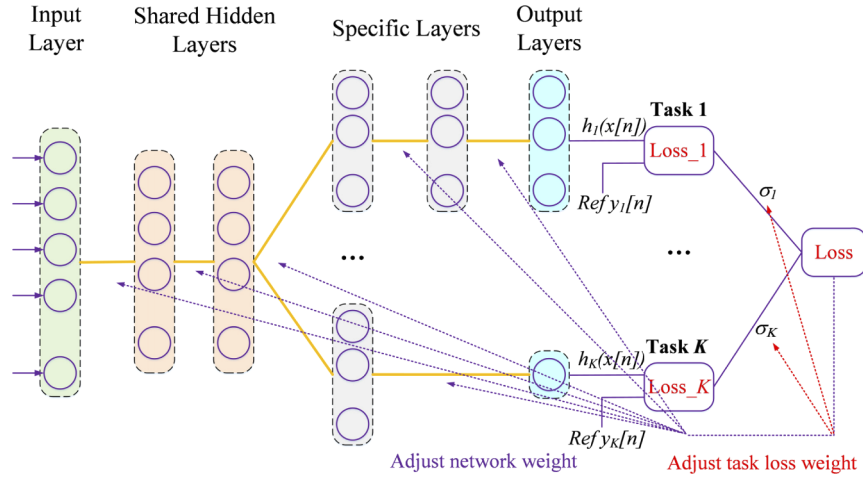


Fig. 4. Schematic structure of loss weight adaptive MTL-ANN.

In Bayesian modelling, homoscedastic uncertainty stays constant for all input data and varies among different tasks which can be described as task-dependent uncertainty [25]. A multi-task loss function based on maximizing the Gaussian likelihood with homoscedastic uncertainty can be derived as following. Let  $f^W(x)$  be the output of a neural network with weights  $W$  on input  $x$ ,  $y$  be the reference output. For regression task, the likelihood can be defined as a Gaussian distribution with a noise scalar  $\sigma$  as shown in Eq. (1). For classification task, the likelihood of model output is squashed to a softmax function with a scaled parameter  $\sigma$  [20]. By adopting  $\sigma$ , classification task and regression task suffer the same scale.

$$p(y|f^W(x)) = N(f^W(x), \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - f^W(x))^2}{2\sigma^2}\right) \quad (1)$$

$$p(y = c|f^W(x), \sigma) = \text{Softmax}\left(\frac{1}{\sigma^2}f^W(x)\right) = \frac{\exp\left(\frac{1}{\sigma^2}f_c^W(x)\right)}{\sum_{c'} \exp\left(\frac{1}{\sigma^2}f_{c'}^W(x)\right)} \quad (2)$$

In the case of multiple model outputs, each task is regarded as statistic independent, so the likelihood is defined to factorise over the outputs as Eq. (3) shows. Where  $y_1, \dots, y_K$  represents reference output of different tasks. Our aim is to maximum the likelihood inference. If we define the loss as Eq. (4), our aim is lead to minimize the loss.

$$p(y_1, \dots, y_K|f^W(x)) = p(y_1|f^W(x)) \cdot \dots \cdot p(y_K|f^W(x)) \quad (3)$$

$$L(W) = -\ln p(y|f^W(x)) \quad (4)$$

Assuming a model's multiple outputs are composed of a classification output  $y_1$  and a regression output  $y_2$ , modelled with a softmax likelihood and a Gaussian likelihood, respectively. The joint



loss is given as:

$$\begin{aligned}
 L(W, \sigma_1, \sigma_2) &= -\ln p(y_1 = c, y_2 | f^W(x)) \\
 &= -\ln p(y_1 = c | f^W(x)) - \ln p(y_2 | f^W(x)) \\
 &\propto -\frac{1}{\sigma_1^2} f_c^W(x) + \ln \left( \sum_{c'} \exp \left( \frac{1}{\sigma_1^2} f_{c'}^W(x) \right) \right) + \frac{1}{2\sigma_2^2} \|y_2 - f^W(x)\|^2 + \ln(\sigma_2)
 \end{aligned} \tag{5}$$

If we write  $L_1(W) = -\ln[\text{Softmax}(y_1, f^W(x))]$  with  $f^W(x)$  not scaled by  $\sigma_1$  for the cross entropy loss of  $y_1$ ,  $L_2(W) = \|y_2 - f^W(x)\|^2$  for the Euclidean loss of  $y_2$ . The joint loss can be rewritten as:

$$\begin{aligned}
 L(W, \sigma_1, \sigma_2) &= \frac{1}{\sigma_1^2} L_1(W) + \frac{1}{2\sigma_2^2} L_2(W) + \ln(\sigma_2) \\
 &\quad - \frac{1}{\sigma_1^2} \ln \left( \sum_{c'} \exp(f_{c'}^W(x)) \right) + \ln \left( \sum_{c'} \exp \left( \frac{1}{\sigma_1^2} f_{c'}^W(x) \right) \right) \\
 &= \frac{1}{\sigma_1^2} L_1(W) + \frac{1}{2\sigma_2^2} L_2(W) + \ln(\sigma_2) + \ln \left( \frac{\sum_{c'} \exp \left( \frac{1}{\sigma_1^2} f_{c'}^W(x) \right)}{\left( \sum_{c'} \exp(f_{c'}^W(x)) \right)^{\frac{1}{\sigma_1^2}}} \right) \\
 &\approx \frac{1}{\sigma_1^2} L_1(W) + \frac{1}{2\sigma_2^2} L_2(W) + \ln(\sigma_1) + \ln(\sigma_2)
 \end{aligned} \tag{6}$$

In the last transition, the last term roughly equal to  $\ln(\sigma_1)$  when  $\sigma_1$  converge to 1. This has the advantage to simplify the optimization objectives, as well as unified the loss form for classification task and regression task. In a word, the general expression of joint loss with homoscedastic uncertainty for MTL based network is:

$$L(W) = \sum_{i=1}^K \left( \frac{1}{\sigma_i^2} L_i(W) + \ln(\sigma_i^2) \right) \tag{7}$$

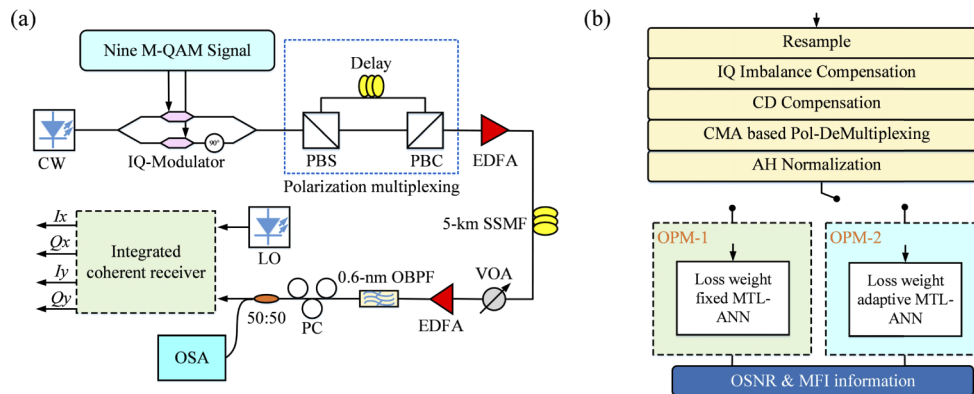
Where  $K$  represents the number of tasks,  $L_i(W)$  represents the loss function of task  $i$ . For the unity of expression, coefficients of the loss function are set as  $1/\sigma_i^2$  for both classification task and regression task.  $\ln(\sigma_i^2)$  can be regarded as regularization term which means the loss will be penalized when setting  $\sigma_i$  too large. Since the loss is smoothly differentiable, Adam algorithm is adopted to adjust network weights and tasks' loss weights in MTL-ANN jointly during the training stage. Compared with traditional stochastic gradient descent algorithm, Adam algorithm is computationally efficient and requires little memory [26].

### 3. Experimental setup and results

#### 3.1. Experimental setup, data collection and network design

The experimental setup of coherent PDM system is shown in Fig. 5(a). An optical carrier at 1552.52 nm generated from a laser with 100 kHz linewidth is injected to the IQ modulator (Fujitsu, FTM7961EX/301). A 50 GSa/s arbitrary waveform generator (AWG, Tektronix AWG70002A) generates the proposed nine modulation format adaptive M-QAM signals with a pattern length of  $2^{13}-1$  symbols at 12.5 GBaud. Signals with peak-to-peak voltage of about 1.2 V drives the IQ modulator. Polarization multiplexing is emulated by utilizing polarization beam splitter (PBS), polarization beam combiner (PBC) and optical delay lines. After that, the generated optical M-QAM signals with  $-2$  dBm power are launched to 5 km standard single mode fiber (SSMF).

After the fiber transmission, a variable optical attenuator (VOA) and an erbium-doped fiber amplifier (EDFA) are employed to load optical noise and adjust OSNR from 10 dB to 25 dB for QPSK, 6-QAM, 8-QAM, and 12-QAM signals, 15 dB to 30 dB for 16-QAM and 24-QAM signals and 20 dB to 35 dB for 32-QAM, 48-QAM, and 64-QAM signals at a step of 1 dB. At the receiver, the signals are passed through a 0.6-nm optical band pass filter (OBPF) and the resulting OSNR is measured by an optical spectrum analyser (OSA). The local oscillator (LO) laser has a linewidth of 100 kHz and its frequency offset with respect to transmitter laser is about 1 GHz. After detected by a 100G dual-polarization integrated coherent receiver (Fujitsu, FTM24706/301), the electrical signals are sampled by a 100 GSa/s digital phosphor oscilloscope (DPO, Tektronix DPO72504D). Finally, the digital signals are processed by the offline DSP as shown in Fig. 5(b).



**Fig. 5.** (a) Experimental setup of coherent PDM system with nine modulation format adaptive M-QAM signals, (b) DSP configuration with two proposed OPM.

At the beginning of the offline DSP flow, the data stream is firstly resampled to enable the proposed algorithm. Then, modulation format independent IQ imbalance compensation and CD compensation are used. Next, we employ CMA-based equalization to de-multiplex PDM signals and compensate linear transmission impairments. After that, circular constellation diagrams with phase rotation are transformed to AHs with 200 bins. Curve fitting is adopted for AHs as shown in Fig. 2(b). Finally, two OPM schemes are deployed for OSNR monitoring and MFI. In OPM-1, a loss weight fixed MTL-ANN is deployed. In OPM-2, our proposed loss weight adaptive MTL-ANN is deployed. Both ANNs have same network structure for comparison. The obtained OSNR information can be used to evaluate the quality of received optical signals. On the other hand, MFI information can be exploited by subsequent modulation format dependent equalization algorithm, like decision-directed least mean square (DD-LMS) algorithm or multi-modulus algorithm (MMA). In this paper, Keras library combined with Tensorflow backend are selected to build the model of ANN [27].

Based on the above system, 100 AHs for each OSNR value of each modulation format (dual-polarization) are collected. The OSNR range is 10-25 dB for QPSK, 6-QAM, 8-QAM, and 12-QAM signal, 15-30 dB for 16-QAM and 24-QAM signal, 20-35 dB for 32-QAM, 48-QAM and 64-QAM signal. So the entire data set comprises 14400 ( $100 \times 16 \times 9$ ) AHs in total. We randomly select 90% (i.e. 12960) AHs in the data set as training set and the remaining 10% (i.e. 1440) AHs are selected as testing set. In our work, validation set is deployed to avoid overfitting problem. 10% of the data in the training set (i.e. 1296) are selected as validation set.

After experimental setup and data collection, ANN structure is investigated. Since the bin number of AHs is set as 200, the neuron number in input layer is 200. The designed ANN contains one shared hidden layer. Two specific hidden layers and one specific hidden layer are

designed for OSNR monitoring and MFI, respectively. The neurons in specific hidden layers are designed a half of the neurons in their front layer and the optimal neuron number in shared hidden layer will be investigated in Section 3.2. Tanh-sigmoid function is selected as activation function for neurons in hidden layers. In this paper, variable learning rate with initial rate of  $1 \times 10^{-3}$  and final rate of  $1 \times 10^{-4}$  is deployed to accelerate network convergence speed and improve performance at same time. Learning rate change step depends on the epochs in Eq. (8), where  $lr$  represents learning rate.

$$lr_{decay} = \frac{lr_{initial} - lr_{final}}{epochs} \quad (8)$$

### 3.2. Comparison of regression and classification problem for OSNR monitoring

OSNR monitoring is usually regarded as a regression problem. Nevertheless, it can also be treated as a classification problem with a proper classification interval. In this section, we investigate the monitoring performance for both two schemes. For regression problem, linear function is selected as activation function in output layer. For classification problem, softmax function is selected as activation function. In this way, 26 OSNR values (from 10 dB to 35 dB with a step of 1 dB) require a one-hot vector with 26 elements and nine modulation formats require a one-hot vector with 9 elements. One-hot vector means a single non-zero vector whose location signifies the true value.

At first, we investigate the MFI accuracy when OSNR monitoring is treated as regression problem and classification problem, respectively. Table 1 shows the MFI accuracy for different modulation formats. The MFI accuracy of nine modulation formats is 100% in the estimated shared hidden layer neuron number range (i.e. 40 to 500) for loss weight adaptive MTL-ANN (i.e. OPM-2) and loss weight fixed MTL-ANN (i.e. OPM-1) with optimal loss ratio. In other word, as to the MTL-ANN, regarding OSNR monitoring as regression problem or classification problem will not affect MFI accuracy.

**Table 1. MFI accuracy for different modulation formats**

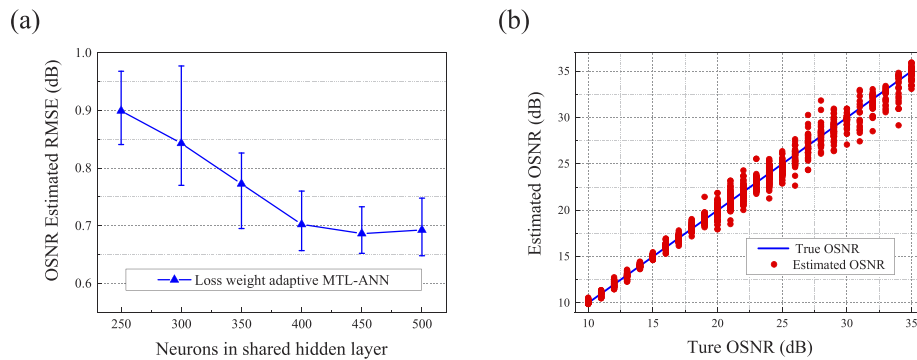
		Identified modulation format								
		QPSK	6QAM	8QAM	12QAM	16QAM	24QAM	32QAM	48QAM	64QAM
Actual modulation format	QPSK	100%	0%	0%	0%	0%	0%	0%	0%	0%
	6QAM	0%	100%	0%	0%	0%	0%	0%	0%	0%
	8QAM	0%	0%	100%	0%	0%	0%	0%	0%	0%
	12QAM	0%	0%	0%	100%	0%	0%	0%	0%	0%
	16QAM	0%	0%	0%	0%	100%	0%	0%	0%	0%
	24QAM	0%	0%	0%	0%	0%	100%	0%	0%	0%
	32QAM	0%	0%	0%	0%	0%	0%	100%	0%	0%
	48QAM	0%	0%	0%	0%	0%	0%	0%	100%	0%
	64QAM	0%	0%	0%	0%	0%	0%	0%	0%	100%

Then we investigate OSNR monitoring performance when treated as a regression problem. Epochs are set big enough (i.e. epochs = 600) to guarantee that the network have reached optimal performance. Since the performance of ANN is affected by the random initialization of ANN weights [11], we evaluate the performance by taking average value, maximum value and minimum value from five random initialization. Figure 6(a) shows the OSNR estimated RMSE versus neurons in shared hidden layer for loss weight adaptive MTL-ANN. The definition of RMSE is shown in Eq. (9). Where  $f(x[n])$  represents the output of ANN for  $x[n]$  and  $y[n]$  represents the corresponding reference value.  $N$  is the number of test data.

$$RMSE(f(x[n]), y[n]) = \sqrt{\frac{1}{N} \sum_{n=1}^N \{f(x[n]) - y[n]\}^2} \quad (9)$$



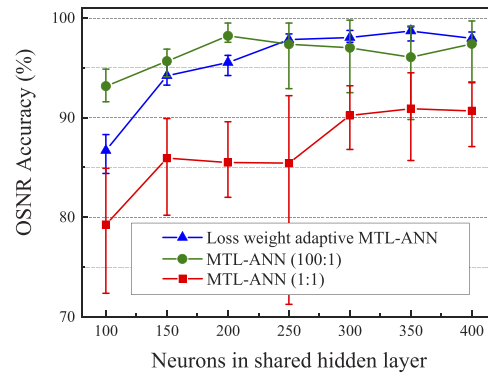
In our experiment, the optimal neuron number in shared hidden layer for weight adaptive MTL-ANN is about 450 and the OSNR estimated RMSE is 0.68 dB at this situation. Figure 6(b) shows the estimated OSNRs versus true OSNRs at optimal network hyperparameters. As shown in the figure, OSNR estimation suffers large estimating error in high OSNR range. To compare the performance of regarding OSNR estimation as regression problem and classification problem, we also investigate the equivalent estimating accuracy of regression problem and equivalent RMSE of classification problem. We treat OSNR estimation as a correct estimation when the estimating error is less than a given threshold. The OSNR estimating accuracy is about 67.8% when the threshold is set as 0.5 dB. Which means 67.8% estimating results are exist within  $[-0.5, 0.5]$  dB estimating deviation.



**Fig. 6.** (a) OSNR estimated RMSE versus neurons in shared hidden layer for loss weight adaptive MTL-ANN, (b) True OSNRs versus estimated OSNRs of loss weight adaptive MTL-ANN (Average, maximum and minimum value from five random initialization).

Since OSNR value has a certain range for signal transmission in realistic field and OSNR monitoring with a proper classification interval can be set for different situation, we also investigate the performance of regarding OSNR monitoring as a classification problem. After optimizing the fixed loss weight ratio, we find the optimal loss ratio of OSNR to MFI is 100:1 for the experimental data. Figure 7 shows the OSNR accuracy versus hidden neurons in shared hidden layer for loss weight fixed and adaptive MTL-ANN. Loss weight adaptive MTL-ANN outperforms loss weight fixed MTL-ANN with 1:1 ratio obviously. More than 7% OSNR accuracy improvement is achieved. Compared with loss weight fixed MTL-ANN with 100:1 ratio, both of them can achieve high OSNR accuracy. In our experiment, the optimal neuron number in shared hidden layer for the optimal weight fixed MTL-ANN is about 200 and the OSNR accuracy is 98.5% at this situation. As for the loss weight adaptive MTL-ANN, the optimal neuron number in shared hidden layer is about 350 and the OSNR accuracy is 98.7% at this situation. The equivalent RMSE of the loss weight adaptive MTL-ANN is 0.66 dB. Note that, while investigating the equivalent RMSE of classification problem, the one-hot vector is firstly transformed to the OSNR value and then RMSE is calculated according to Eq. (9).

The comparison of regarding OSNR monitoring as regression and classification problem is summarized in Table 2. Compared with regarding OSNR monitoring as a regression problem, regarding it as a classification problem achieves similar RMSE and much higher accuracy. This means the inaccurately estimated OSNR suffers large vibration for classification problem. However, the inaccurately estimated OSNR is only a small part and can be identified by analyzing the OSNR output vector of neural network. Therefore, regarding OSNR monitoring as a classification problem achieves better OSNR monitoring performance. OSNR monitoring is selected as a classification problem in the following part.



**Fig. 7.** OSNR accuracy versus hidden neurons in shared hidden layer for loss weight fixed and adaptive MTL-ANN (Average, maximum and minimum value from five random initialization).

**Table 2. Comparison of regression and classification problem for OSNR monitoring**

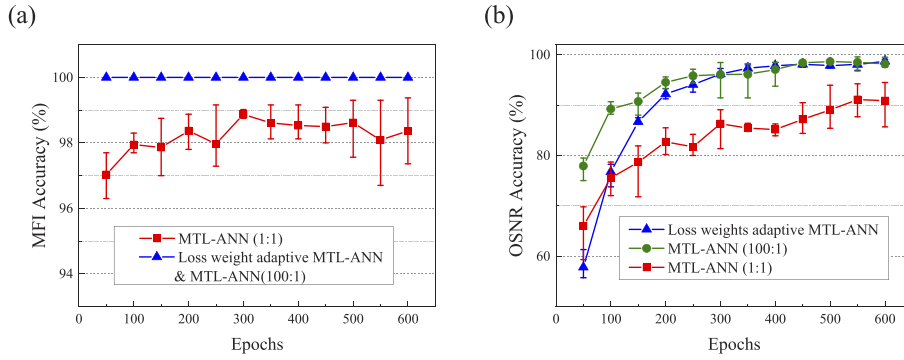
OSNR monitoring problem type	RMSE (dB)	Accuracy (%)
Regression	0.68	67.8
Classification	0.66	98.7

### 3.3. Experimental results and discussions

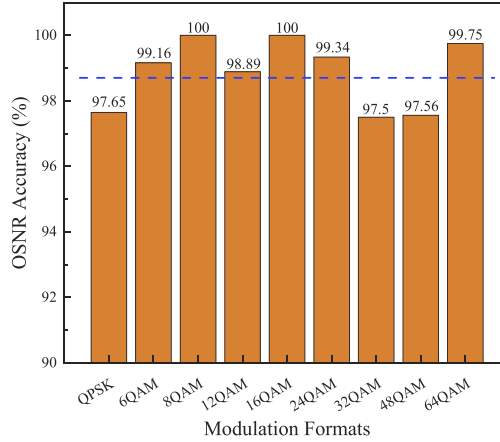
In this section, OSNR monitoring is selected as a classification problem. Figure 8(a) shows the MFI accuracy versus epochs at optimal neuron number in shared hidden layer. For loss weight adaptive MTL-ANN and loss weight fixed MTL-ANN with 100:1 loss ratio, the MFI accuracy is 100% when the number of epoch is larger than 50. As for loss weight fixed MTL-ANN with 1:1 ratio, MFI accuracy roughly rises under 300 epochs, which means 300 epochs are enough for MFI task. OSNR estimating accuracy versus epochs is shown in Fig. 8(b). Compared with MFI task, more epochs are needed for OSNR monitoring task, which means OSNR monitoring is more difficult. The conclusion is same as [11]. For loss weight adaptive MTL-ANN, OSNR accuracy improves rapidly with the increase of epochs and a floor appears when epochs reach 400. As for loss weight fixed MTL-ANN with 100:1 ratio, it outperforms adaptive MTL-ANN when the number of epoch is less than 300. About 450 epochs are needed to achieve the optimal accuracy. As for loss weight fixed MTL-ANN with 1:1 ratio, 550 epochs are needed, which indicates training time can be reduced by choosing loss weight ratio properly. Note that, the optimal OSNR estimating accuracy is 98.7% for loss weight adaptive MTL-ANN.

After investigating the OSNR estimating accuracy for the group of modulation formats, the average OSNR estimating accuracy from five random initialization for different modulation format is shown in Fig. 9. The blue dotted line represents the average optimal OSNR accuracy of all modulation formats (i.e. 98.7%). The OSNR accuracy is higher than 97.5% for all modulation formats, which means the proposed method is effective for all modulation formats under consideration. Except QPSK, 32QAM and 48QAM, the OSNR accuracy of other six formats is higher than 98.7%.

Table 3 shows the optimal loss weight ratio for loss weight adaptive MTL-ANN in random initialization. Different with loss weight fixed MTL-ANN, which only adjust network weights during the training stage, the proposed adaptive method adjust tasks' loss weights and network weights jointly. Therefore, the optimal loss weight ratio found by the proposed adaptive method varies with different network weights. In other word, the optimal ratio varies with random initialization of network. As can be seen from Table 3, the similar OSNR accuracy can be



**Fig. 8.** (a) MFI accuracy and (b) OSNR accuracy versus epochs for loss weight fixed and adaptive MTL-ANN (Average, maximum and minimum value from five random initialization).



**Fig. 9.** OSNR accuracy for different modulation format in loss weight adaptive MTL-ANN (Average value from five random initialization).

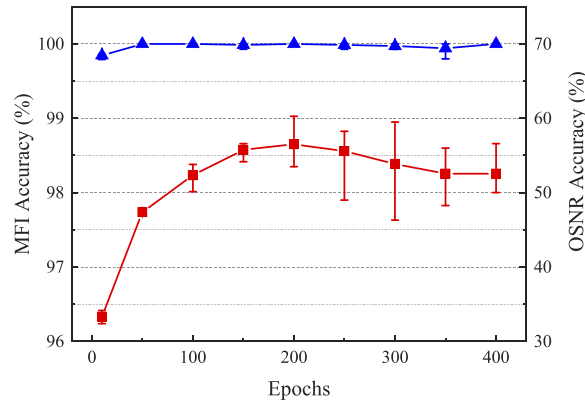
achieved with different loss weight ratio. This means the optimal loss weight ratio for the proposed adaptive method is not fixed for a specific link configuration and several loss weight ratios can achieve the optimal performance.

**Table 3.** Loss weight ratio for loss weight adaptive MTL-ANN in random initialization

	Random initialization				
	1	2	3	4	5
Loss weight ratio (MFI to OSNR)	2201	2444	637	559	1465
OSNR accuracy (%)	98.3	97.7	97.5	99.1	98.75

Finally, we investigate the performance of loss weight adaptive MTL-ANN without the regularization term mentioned in Eq. (7). As shown in Fig. 10, MFI accuracy can reach 99.8% without regularization term since it is a simple task. As for OSNR monitoring, the accuracy is about 55%. Compared with loss weight adaptive MTL-ANN with regularization term, the OSNR accuracy decreases more than 40%. The reason is that directly learning the loss weight

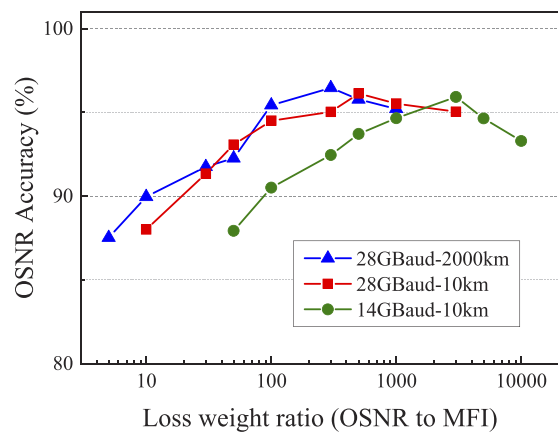
without regularization term will result in fast convergence to zero for joint loss [20]. Therefore, the regularization terms are indispensable for difficult tasks in the weight adaptive MTL method.



**Fig. 10.** MFI accuracy and OSNR accuracy versus epochs for loss weight adaptive MTL-ANN without regularization term (Average, maximum and minimum value from five random initialization).

#### 4. Discussion of loss weight adaptive MTL-ANN in simulation system

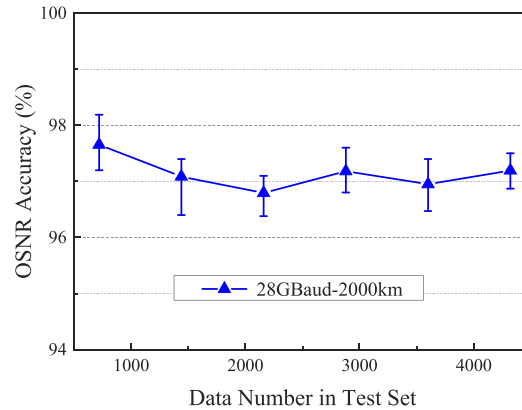
In this part, we build a simulation system based on VPI transmission Maker 9.1 to investigate the necessity and generalization ability of the loss weight adaptive MTL-ANN. As discussed in Section 3, loss weight adaptive MTL-ANN and loss weight fixed MTL-ANN with optimal loss weight ratio both achieve similar MFI and OSNR monitoring performance. Besides, fewer neurons and same epochs are needed for loss weight fixed MTL-ANN with optimal loss weight ratio. However, the optimal loss weight ratio varies with different link configurations. Figure 11 shows OSNR accuracy versus loss weight ratio for three link configurations in simulation system. As can be seen from the figure, the optimal loss weight ratio for the proposed nine modulation format adaptive M-QAM at 28 GBaud after 2000 km transmission is 300. As for 28 GBaud signals after 10 km transmission, the optimal ratio is 500. As for 14 GBaud signals after 10 km



**Fig. 11.** OSNR accuracy versus loss weight ratio (OSNR to MFI) for loss weight fixed MTL-ANN in simulation.

transmission, the optimal ratio is 3000. Therefore, it is time-consuming and highly complex to find the optimal loss weight ratio for different link configurations manually.

Figure 12 shows the OSNR accuracy versus the data number in test set. Data obtained from simulation system with nine M-QAM signals at 28 GBaud after 2000 km transmission are used to test the performance of OSNR estimation. 10080 ( $70 \times 16 \times 9$ ) AHs are collected for train set. As can be seen from Fig. 12, OSNR estimation has a stable performance when the test data set is obtained from a longer duration, which means the proposed loss weight adaptive MTL-ANN has an excellent generalization ability.



**Fig. 12.** OSNR accuracy versus data number in test set for loss weight adaptive MTL-ANN in simulation (Average, maximum and minimum value from five random initialization).

## 5. Conclusion

In this paper, we have applied a loss weight adaptive MTL-ANN based optical performance monitor to simultaneously monitor OSNR and identify modulation formats with signals' AHs as input features. An experimental setup for coherent PDM system with 5 km SSMF transmission is deployed to evaluate the performance of the proposed monitor. The experimental results show that the MFI accuracy of nine adaptive M-QAM considered reaches 100% in the estimated OSNR range. Besides, OSNR estimation with RMSE of 0.68 dB and accuracy of 98.7% are achieved when treated as regression problem and classification problem, respectively. The OSNR accuracy for all modulation formats is higher than 97.5%. We also investigate the equivalent OSNR estimating accuracy and RMSE for regression problem and classification problem, respectively. After comparison, we find regarding OSNR monitoring as a classification problem achieves better OSNR monitoring performance.

Through investigating the optimal loss weight ratio in random initialization, we find several loss weight ratios can achieve the optimal performance for the proposed adaptive method. The importance of regularization terms in joint loss function for MTL-ANN is also investigated. Furthermore, in the simulation system, we demonstrate that the proposed loss weight adaptive MTL-ANN has an excellent generalization ability when the test data is obtained from a longer duration. Though there is almost no performance difference between loss weight adaptive MTL-ANN and loss weight fixed MTL-ANN with optimal loss weight ratio, the proposed loss weight adaptive MTL-ANN can search the optimal loss weight ratio automatically. This feature is time-saving and low complexity since the optimal loss weight ratio varies for different link configuration. By adopting this loss weight adaptive method, the number of estimated parameters can be easily expanded, which is attractive for real-time multiple parameters estimation in future heterogeneous optical network.



## Funding

National Natural Science Foundation of China (61431003, 61601049, 61625104, 61821001, 61901045); Key Technologies Research and Development Program (2018YFB2201803); Fundamental Research Funds for the Central Universities; BUPT Excellent Ph.D. Students Foundation (CX2018115); Beijing Municipal Science and Technology Commission (Z181100008918011); State Key Laboratory of Information Photonics and Optical Communications (IPOC2017ZT08).

## Disclosures

The authors declare no conflicts of interest.

## References

1. Cisco, "Cisco visual networking index: forecast and trends, 2017-2022," (Cisco White Paper, 2019). <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf>.
2. K. Roberts, Q. Zhuge, I. Monga, S. Gareau, and C. Laperle, "Beyond 100 Gb/s: Capacity, Flexibility, and Network Optimization," *J. Opt. Commun. Netw.* **9**(4), C12–C24 (2017).
3. Q. Zhuge and W. Hu, "Application of Machine Learning in Elastic Optical Networks," in *2018 European Conference on Optical Communication (ECOC)* (IEEE, 2018), pp. 1–3.
4. A. Yi, L. Yan, H. Liu, L. Jiang, Y. Pan, B. Luo, and W. Pan, "Modulation format identification and OSNR monitoring using density distributions in Stokes axes for digital coherent receivers," *Opt. Express* **27**(4), 4471 (2019).
5. 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).
6. W. Freude, R. Schmogrow, B. Nebendahl, M. Winter, A. Josten, D. Hillerkuss, S. Koenig, J. Meyer, M. Dreschmann, M. Huebner, C. Koos, J. Becker, and J. Leuthold, "Quality metrics for optical signals: Eye diagram, Q-factor, OSNR, EVM and BER," in *2012 14th International Conference on Transparent Optical Networks (ICTON)* (2012), pp. 1–4.
7. I. Tomkos, S. Azodolmolky, J. Solé-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).
8. S. M. Bilal, G. Bosco, Z. Dong, A. P. T. Lau, and C. Lu, "Blind modulation format identification for digital coherent receivers," *Opt. Express* **23**(20), 26769 (2015).
9. 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 MLSD," *J. Lightwave Technol.* **35**(23), 5193–5200 (2017).
10. 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 (2017).
11. Z. Wan, Z. Yu, L. Shu, Y. Zhao, H. Zhang, and K. Xu, "Intelligent optical performance monitor using multi-task learning based artificial neural network," *Opt. Express* **27**(8), 11281 (2019).
12. D. Wang, M. Wang, M. Zhang, Z. Zhang, H. Yang, J. Li, J. Li, and X. Chen, "Cost-effective and data size-adaptive OPM at intermediated node using convolutional neural network-based image processor," *Opt. Express* **27**(7), 9403 (2019).
13. 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 (2015).
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 (2018).
15. 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 (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 (2018).
17. L. Baker-Mefflah, 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).
18. 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).
19. 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 (2018).
20. R. Cipolla, Y. Gal, and A. Kendall, "Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 7482–7491.

21. A. T. Le and K. Araki, "A group of modulation schemes for adaptive modulation," in *2008 11th IEEE Singapore International Conference on Communication Systems* (IEEE, 2008), pp. 864–869.
22. Z. Yu, H. Chen, M. Chen, S. Yang, and S. Xie, "Bandwidth Improvement Using Adaptive Loading Scheme in Optical Direct-Detection OFDM," *IEEE J. Quantum Electron.* **52**(10), 1–6 (2016).
23. R. A. Caruana, "Multitask Learning: A Knowledge-Based Source of Inductive Bias," in *Machine Learning Proceedings 1993* (Elsevier, 1993), pp. 41–48.
24. M. Long, Z. Cao, J. Wang, and P. S. Yu, "Learning Multiple Tasks with Multilinear Relationship Networks," arXiv:1506.02117 [cs] (2015).
25. A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" arXiv:1703.04977 [cs] (2017).
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).