IEEE OSA The Optical Society | Journal of Lightwave Technology

Low complexity OSNR monitoring and modulation format identification based on binarized neural networks

Journal:	*Journal of Lightwave Technology*
Manuscript ID	JLT-24619-2019.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	22-Sep-2019
Complete List of Authors:	Zhao, Yilun; Beijing University of Posts and Telecommunications, State Key Laboratory of Information Photonics and Optical Communications Yu, Zhenming; Beijing University of Posts and Telecommunications, State Key Laboratory of Information Photonics and Optical Communications Wan, Zhiquan; Beijing University of Posts and Telecommunications, State Key Laboratory of Information Photonics and Optical Communications Hu, Shaohua; University of Electronic Science and Technology of China, School of Information and Communication Engineering Shu, Liang; Beijing University of Posts and Telecommunications, State Key Laboratory of Information Photonics and Optical Communications Zhang, Jing; University of Electronic Science and Technology of China, School of Information and Communication Engineering Xu, Kun; Beijing University of Posts and Telecommunications, State Key Laboratory of Information Photonics and Optical Communications
Key Words:	Modulation format identification, OSNR monitoring, Binarized neural networks, Optical communications
Category:	Systems and Subsystems

Low complexity OSNR monitoring and modulation format identification based on binarized neural networks

Yilun Zhao, Zhenming Yu, Zhiquan Wan, Shaohua Hu, Liang Shu, Jing Zhang, and Kun Xu, *Member, IEEE*

***Abstract*—We propose and experimentally demonstrate a method of optical signal-to-noise ratio (OSNR) monitoring and modulation format identification (MFI) using a binarized convolutional neural network (B-CNN) in coherent receiver. The proposed technique automatically extracts OSNR and modulation format dependent features from the signals' ring constellation maps. The experimental results show that the MFI accuracy can reach 100% for the common nine modulation formats. The OSNR monitoring accuracy can reach 95.83% for 64QAM signal and 99.38% for the other eight modulation formats, respectively. The memory consumption of B-CNN is less than 1/3 of the float value convolutional neural network (F-CNN), which implies that the computational efficiency can be improved significantly without performance loss. Therefore, it is attractive for cost-effective multi-parameter estimation in next-generation optical networks.**

***Index Terms*—Modulation format identification (MFI), OSNR monitoring, binarized neural networks (BNNs), optical fiber communications.**

I. Introduction

With the increasing of the demand for various data services, optical network is evolving from current fixed architecture to future flexible one [1], [2]. In order to provide agile and affordable on-demand network service, it is desirable to develop a cognitive optical network (CON). The aim of CON is to introduce intelligence into the control plane that allows for autonomous end-to-end performance optimization and minimization of required human intervention [3]. A CON is also capable of coordinating transmitters and receivers for real-time dynamic adjustment of the modulation format, line rate, spectrum allocation and other parameters. Managing the transceivers in this way could help to realize intelligent channel management, bandwidth allocation and other link functions at the network nodes, thereby improving service quality and transmission quality [4]. In order to reduce operating costs, ensure optimum resources utilization and guarantee adequate operation and management of dynamic optical networks, it is essential to continuously monitor various network performance parameters, which is referred to as optical performance monitoring (OPM) [5]. Since most of the linear impairments can be fully compensated, transmission performance is largely determined by the OSNR and hence OSNR monitoring is especially vital for coherent links. Moreover, the OSNR monitoring technique must be appropriate for the incoming signal type, and the carrier recovery module used in the receiver must be suitable for the received modulation format. Therefore, MFI is also indispensable for digital receivers in CON.

Recently, the combination of OSNR estimation and MFI is becoming a trend of development. Machine learning (ML) algorithms offer powerful tools to solve various problems in many areas and are widely explored in optical communication systems [6]–[8]. Some researches have employed ML techniques to achieve joint OSNR monitoring and MFI [9]–[14]. In [9], fully-connected neural networks are used for OSNR monitoring and MFI in digital coherent receivers by using amplitude histograms (AHs) as input features. Meanwhile, the concept of intelligent constellation diagram analyzer were proposed in [10] to achieve MFI and OSNR monitoring simultaneously by using convolutional neural network (CNN). The authors have proved that CNN outperforms other ML algorithms such as decision tree, multi-layer processor (MLP), support vector machine (SVM) and k-nearest neighbors (KNN). Recently, multi-task learning based schemes are proposed to achieve joint OSNR monitoring and MFI using one single neural network [11]–[13]. In addition, a CNN based OPM is proposed at intermediate node by analyzing asynchronous delay-tap plots (ADTPs) [14]. Although neural network has a strong ability in joint MFI and OSNR monitoring, existing schemes are computationally expensive and memory intensive, which is not proper for the deployment in receivers with low memory resources and strict latency requirements [15].

In this paper, we propose a novel intelligent optical performance monitoring technique based on binarized convolutional neural network (B-CNN), which can achieve high accuracies of joint MFI and OSNR monitoring with ultralow memory consumptions and computation complexity. The B-CNN is constructed based on Binarized Neural Networks (BNNs) [16] with binary weights and activations at run-time during the forward pass, BNNs drastically reduce memory size and accesses, and is expected to substantially improve power-efficiency. We set up our simulation in two stages. In the first

This work was supported in part by the NSFC Program (No. 61431003, 61601049, 61625104, 61821001, 61420106011); Fundamental Research Funds for the Central Universities; Beijing Municipa Science & Technology Commission (Grant No. Z181100008918011); Fund of State Key Laboratory of Information Photonics and Optical Communications, BUPT (No. IPOC2017ZT08). (*Corresponding author: Zhenming Yu*)

Y. Zhao, Z. Yu, Z. Wan, L. Shu, and K. Xu are with the State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing, 100876, China (e-mail: zyl2018111137@bupt.edu.cn; yuzhenming@bupt.edu.cn; wanzhiquan@bupt.edu.cn; shuliang@bupt.edu.cn; xukun@bupt.edu.cn).

S. Hu and J. Zhang, was with Key Laboratory of Optical Fiber Sensing and Communications, University of Electronic Science and Technology of China, Chengdu, 611731, China (email:hshhfmy@std.uestc.edu.cn; zhangjing1983@uestc.edu.cn).

stage, we set up the single-polarization (SP) flexible quadrature amplitude modulation (QAM) coherent optical communication system. We verify the effectiveness of B-CNN and compare it with float value convolutional neural network (F-CNN). In the second stage, we set up the dual-polarization (DP) system with different link scenarios and then study the robustness of the proposed method. The results show that B-CNN is able to achieve similar performance with F-CNN with much lower memory consumption and computational complexity, and is robust to different link scenarios. In addition, we further verify our scheme in the experimental setup of 12.5 Gbaud flexible DP-QAM system with transmission of 5 km SSMF. The experimental results show that accuracies of OSNR estimation higher than 95% can be achieved based on B-CNN, and the accuracies of MFI are 100%.

II. Operating Principle

In the proposed B-CNN based optical performance monitor, constellation diagrams after CD compensation (for SP-QAM signals) or after CD compensation and constant modulus algorithm (CMA) (for polarization demultiplexing of DP-QAM signals) are selected as processing objective. In order to reduce the computation load and enhance the ability to express the characteristics of different modulation formats, we directly generate the grayscale maps of constellation diagrams instead of first collect colored constellation diagrams [10], [14] since the grayscale map has only one channel. The value of the pixel represents the density of the constellation points of the corresponding grid. The grayscale maps of the adopted modulation format at OSNR of 30 dB are shown in Fig. 1, maps of higher order modulation formats consist of more rings since their original signals have more different moduli. Finally, the grayscale maps are sent to B-CNN for MFI and OSNR monitoring.

When training a BNN, both the weights and the activations are constrained to either +1 or -1. We select a deterministic function to transform the real-valued variables into those two values:

$$x^b = \mathrm{Sign}(x) = \begin{cases} +1 & \text{if } x \geq 0, \\ -1 & \text{otherwise}, \end{cases} \tag{1}$$

where x^b is the binarized variable (weight or activation) and x is the real-valued variable. Since the derivative of the sign function is zero almost everywhere, making it apparently incompatible with backpropagation, the authors in [19] provided an approach which they refer to as straight-through estimator that take into account the saturation effect, and does use the deterministic rather than stochastic sampling of the bit. Consider the *sign* function quantization

$$q = \mathrm{Sign}(r) \tag{2}$$

and assume that an estimator g_q of the gradient $\partial C/\partial q$ has been obtained (with the straight-through estimator when needed). Then, the straight-through estimator of $\partial C/\partial r$ is simply [16]

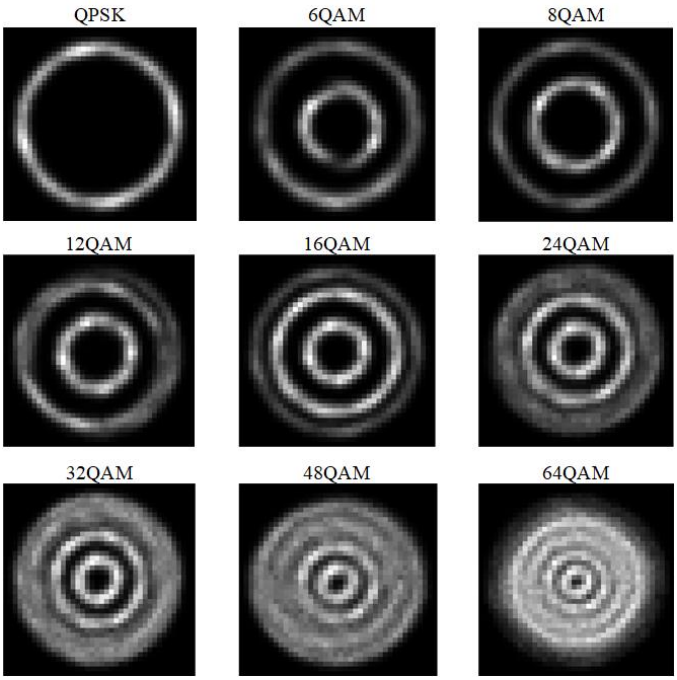

Fig. 1. The grayscale map of each modulation format at OSNR of 30 dB.

$$g_r = g_q \times 1_{|r|\leq 1}. \tag{3}$$

Here C denotes the cost function of the neural network, r is the input of an neuron and q is the output after *sign* function, g_r and g_q are the estimation of the gradient $\partial C/\partial r$ and $\partial C/\partial q$ respectively. Note that this preserves the gradient's information and cancels the gradient when r is too large. To be more specific, when the input is either larger than 1 or smaller than -1, we cancel its gradient and do not update its value anymore, otherwise we assume the gradient is 1 and spread it in the back-propagation process. The use of this straight-through estimator can also be seen as propagating the gradient through *hard tanh*, which is the following piece-wise linear activation function:

$$\mathrm{Htanh}(x) = \mathrm{Clip}(x, -1, 1) = \max(-1, \min(1, x)). \tag{4}$$

For hidden units, we use the sign function non-linearity to obtain binary activations, and for weights we first constrain each real-valued weight between -1 and 1, by projecting w^r outside of $[-1, 1]$, i.e., clipping the weights during training, then we quantize w^r using $w^b = \mathrm{Sign}(w^r)$. Therefore, we can build up B-CNN based on common CNN model by implementing the methods above.

III. Simulation Setup

A. *Optical communication system*

The simulation setup is shown in Fig. 2. We generate optical signals with nine different modulation formats (QPSK, 6QAM, 8QAM, 12QAM, 16QAM, 24QAM, 32QAM, 48QAM, 64QAM) at 28Gbaud. The signals are amplified using an erbium-doped fiber amplifier (EDFA) and sent over

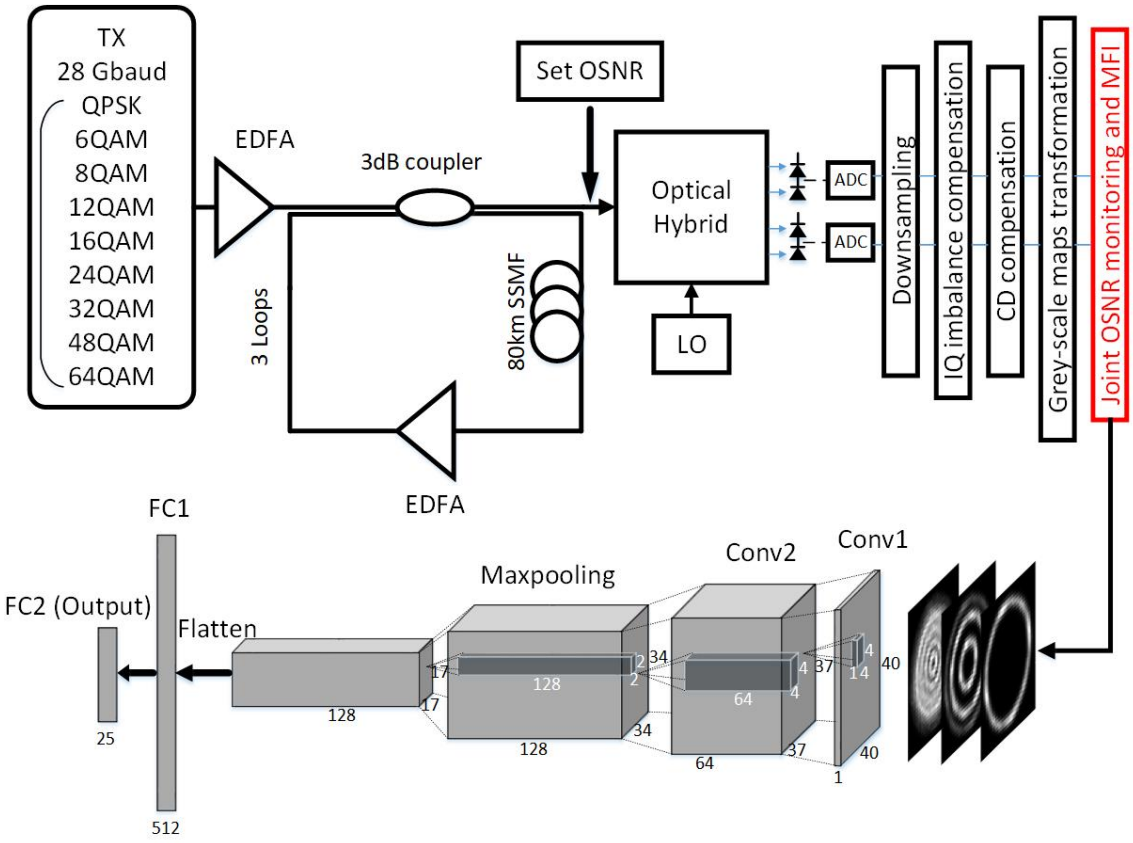

Fig. 2. Simulation setup and B-CNN architecture. EDFA: erbium-doped fiber amplifier; SSMF: standard single mode fiber; LO: local oscillation; ADC: analog-digital converter; Conv: convolutional layer; FC: fully-connected layer.

a fiber recirculating loop consists of an 80 km long standard single-mode fiber (SSMF). An OSNR module is used to adjust OSNR values before the receiver by adding Gaussian white noise to the input signals. At the receiver, the signal is coherently detected by the conventional coherent receiver. After synchronous sampling by two analog-to-digital converts (ADC), two electrical digital signals containing the in-phase (I) and quadrature (Q) information are obtained and sent into offline digital signal processing (DSP) module. As depicted in Fig. 2, after downsampling, IQ imbalance compensation and CD compensation, the received signal is utilized to generate the ring grayscale constellation maps for B-CNN to recognize modulation formats and estimate OSNR values.

B. *Details of B-CNN model*

1) Network structure: We build up the neural network model based on [17]. The network structure of B-CNN is also displayed in Fig. 2, which consists of 2 convolutional layers, 1 maxpooling layer and 2 fully-connected (FC) layers. The first layer has 64 filters (convolution kernels) and the second layer has 128 filters. All the filters have the same size of 4×4. Then the 2 convolutional layers are followed by a 2×2 maxpooling layer to reduce parameters size. Finally, the three-dimensional data is flattened and then processed by 2 FC layers and the first FC layer has 512 neuron nodes. The second FC layer has 25 nodes, which are equal to the number of elements of target label vectors. As a comparison, the F-CNN with the similar architecture is adopted. In F-CNN, the filter numbers for the two convolutional layers are 32 and 64 respectively, the filter size is 3×3. The first FC layer of F-CNN contains 128 neuron nodes and the second FC layer has 25 nodes.

2) Methods to measure the performance of model: For classification tasks, *cross-entropy* loss is commonly used to measure the performance of a model whose output is a probability between 0 and 1. It is calculated as $-\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$, where M is the number of classes, $y_{o,c}$ is a binary indicator (0 or 1) if class label c is correct classification for observation o and $p_{o,c}$ is the predicted probability observation o is of class c. The predicted probability $p_{o,c}$ is calculated by using *softmax* as activation function in the output layer. In B-CNN, since all the activation functions are replaced with *hard-tanh* in training stage and *sign* function is adopted in running stage, there exist negative values in the output of last layer i.e. the output is not a probability between 0 and 1. As a result, the *cross-entropy* loss fail to work. Therefore, we choose *squared-hinge* function as the loss function defined by eq. (5), where y is the output of a neuron and t is the target value. For F-CNN, *relu* function and *sigmoid* function are selected as activation functions for hidden layers and output layer, respectively. In addition, we choose *binary-crossentropy* function as the loss function since it is more suitable for multi-label tasks. For both B-CNN and F-CNN, *Adam* optimizer is adopted because it is computationally efficient and requires little memory [18]. Besides, we adopt *Batchnormalization* between each layer by

TABLE I
EXAMPLES OF LABEL VECTORS

Signal type	Label vector
QPSK, 15 dB	$\underbrace{+1-1\cdots-1}_{9}\underbrace{+1-1\cdots-1}_{16}$
QPSK, 16 dB	$\underbrace{+1-1\cdots-1}_{9}\underbrace{-1+1\cdots-1}_{16}$
QPSK, 30 dB	$\underbrace{+1-1\cdots-1}_{9}\underbrace{-1\cdots-1+1}_{16}$
16QAM, 15 dB	$\underbrace{-1\cdots+1\cdots-1}_{9}\underbrace{+1-1\cdots-1}_{16}$
16QAM, 16 dB	$\underbrace{-1\cdots+1\cdots-1}_{9}\underbrace{-1+1\cdots-1}_{16}$
16QAM, 30 dB	$\underbrace{-1\cdots+1\cdots-1}_{9}\underbrace{-1\cdots-1+1}_{16}$
64QAM, 15 dB	$\underbrace{-1-1\cdots+1}_{9}\underbrace{+1-1\cdots-1}_{16}$
64QAM, 16 dB	$\underbrace{-1-1\cdots+1}_{9}\underbrace{-1+1\cdots-1}_{16}$
64QAM, 30 dB	$\underbrace{-1-1\cdots+1}_{9}\underbrace{-1\cdots-1+1}_{16}$

performing the normalization for each mini-batch to accelerate network training [19].

$$L(y,t) = \max\left(0, (1-yt)^2\right) \tag{5}$$

C. *Data preparation*

Based on the above scheme, we collect 100 grayscale images for each OSNR value of each modulation format as data set. Then we randomly mess up the data set and take 70% as a training set and 30% as the test set. More specifically, each format has 1600 grayscale maps for 16 OSNR values ($15 \sim 30$ dB for QPSK, 6QAM, 8QAM, 12QAM, 16QAM, 24QAM, 32QAM, $20 \sim 35$ dB for 48QAM, and $25 \sim 40$ dB for 64QAM). In the training set, each map has a label vector. In CNN, the label vector is composed of several binary bits, the amount of binary bits is equal to the class number. In our scheme, since all the weights and neurons are binarized and constrained to +1 and -1, the elements of the label vector are either +1 or -1. We set the first 9 bits to denote the modulation formats, and the other 16 bits to denote the 16 OSNR values. Thus the output label vector consists of 25 bits totally. Table I illustrates the relationship between modulation formats and corresponding label vectors by listing the labels of QPSK, 16QAM, and 64QAM signals.

IV. SIMULATION RESULTS AND DISCUSSION

A. *Effectiveness of B-CNN*

We first evaluate the proposed method in SP-QAM system with transmission of 240 km SSMF. Since the performance of neural networks is affected by the random initialization of weights, we evaluate the performance by taking average value from five random initializations. First, we set the resolution of grayscale maps at 40×40, i.e. the pixel size of each map is 40×40 and the values of each pixel are normalized to [0, 255], representing the density of points of the corresponding grid. We accumulate 2×10^5 signal points to generate each grayscale map, i.e. each map contains 2×10^5 points. In this case, the accuracies of MFI and OSNR estimation are both 100%. Then we further investigate the influence of resolution and points number. Fig. 3(a) shows the accuracy of OSNR estimation of each modulation format as a function of the resolutions of grayscale maps. As the resolution increases, the rings of grayscale maps become more clear, thus the accuracies of OSNR estimation increase. With the resolution higher than 30×30, the accuracies of OSNR estimation of nine modulation formats are all 100%. It is evident from the figure that different modulation formats have different tolerances for resolution. In general, higher order modulation formats require higher resolution. When the resolution decreases lower than 20×20, the accuracies of OSNR estimation of 48QAM and 64QAM decrease dramatically. Moreover, we investigated how many signal points should be included in a single constellation diagram. The results shown in Fig. 5(b) indicates that the accuracy of OSNR estimation increases as the number of points increases. In general, higher order modulation formats require more points to achieve better accuracy. Since higher order modulation formats have more rings than lower order modulation formats, they have more complicated constellation diagrams. Therefore, we need more detailed information for higher order modulation formats.

B. *Comparison between B-CNN and F-CNN*

1) Parameters size and computational efficiency: Fig. 4 shows the parameters size of F-CNN and B-CNN. Although both the filters number and the number of the hidden units of B-CNN are larger than that of F-CNN, the total parameters size of B-CNN is much lower. To be more exact, the B-CNN with the architecture described in section III has 3.14 MB parameters, while the F-CNN for comparison has 10.23 MB parameters. Thanks to the binarization of B-CNN, most of its parameters occupy only 1 bit memory, thus indicates 32 time lower memory consumption than 32-bit float value. Then we measure the accuracies of OSNR estimation versus number of epochs for B-CNN and F-CNN simultaneously, where an epoch is a complete process through the whole training set. The result is shown in Fig. 5. We choose the accuracies of OSNR estimation for 64QAM to depict the accuracy as a function of number of epochs. Results show that the B-CNN requires 9 epochs to achieve 100% accuracy and the F-CNN requires 20 epochs. It is obvious that the proposed B-CNN scheme requires less epochs to achieve stable performance, which indicates that B-CNN require less offline training overhead. Therefore, B-CNN will be more efficient when deployed in actual OPM devices.

2) Requirements for input features: Fig. 6, 7 show the accuracies of OSNR estimation for each modulation format by B-CNN and F-CNN with different input features, i.e. grayscale maps with different resolutions and numbers of signal points. Fig. 6 shows that the F-CNN requires lower resolution of grayscale maps to achieve 100% accuracy. In other words,

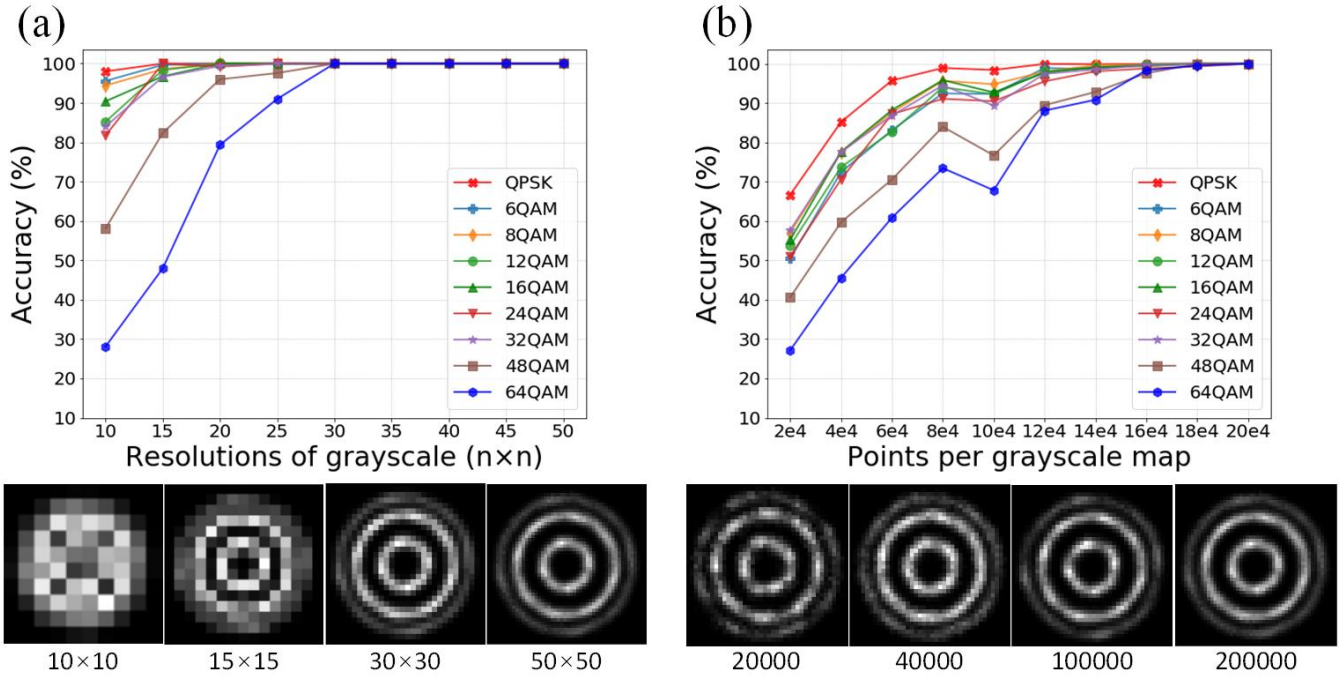

Fig. 3. (a) Accuracy of OSNR estimation vs. resolutions of grayscale, (b) Accuracy of OSNR estimation vs. points per grayscale map.

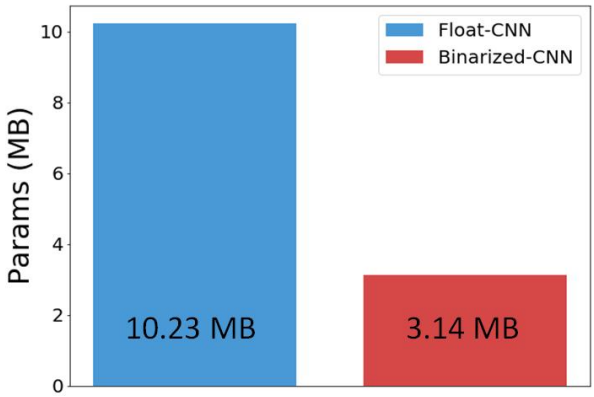

Fig. 4. Comparison of Float-CNN and B-CNN in parameters size.

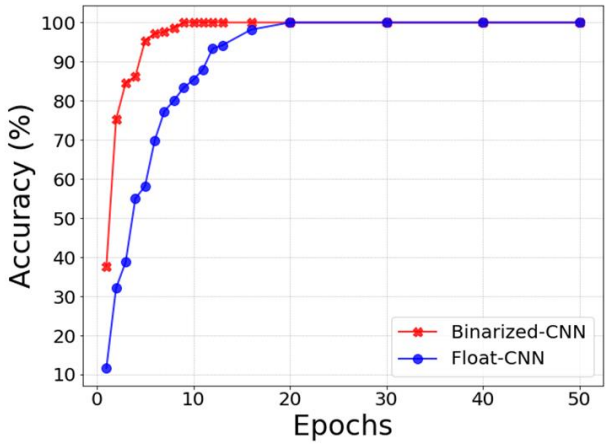

Fig. 5. Accuracy of OSNR estimation in epochs number using B-CNN and F-CNN for 64QAM.

the F-CNN has a higher tolerance of resolution. Similarly, Fig. 7 shows that F-CNN also has a higher tolerance of points number. Nonetheless, when the resolution is higher than 30×30 and the points number is larger than 1.6×10^5, the performances of B-CNN and F-CNN are similar. Compared with OSNR estimation, the task of MFI is much easier to attain 100% accuracy. Fig. 8 shows the MFI accuracy at different resolution and signal points in a map, the accuracy of MFI remains higher than 99.5% either under low resolution or with small amounts of points. As discussed above, the constellation maps of different modulation formats contain different number of rings, which is an obvious feature for classification assisted by ML techniques. For the comparison, we also depict the results of F-CNN in the figure. While the F-CNN is able to attain 100% accuracy of MFI at lower resolution and less points, there is a certain level of performance loss when using B-CNN.

3) Advantages of B-CNN: The simulation results indicate that when supporting large number of modulation formats, the network requires much more input features and become more complicated, which greatly increase the computational complexity in the propagation process. In order to respond quickly to real-time changes in the optical networks, how to reduce the calculation latency is critical to the deployment of deep CNNs. It is possible to speed up GPU implementations of BNNs by SIMD (single instruction, multiple data) within

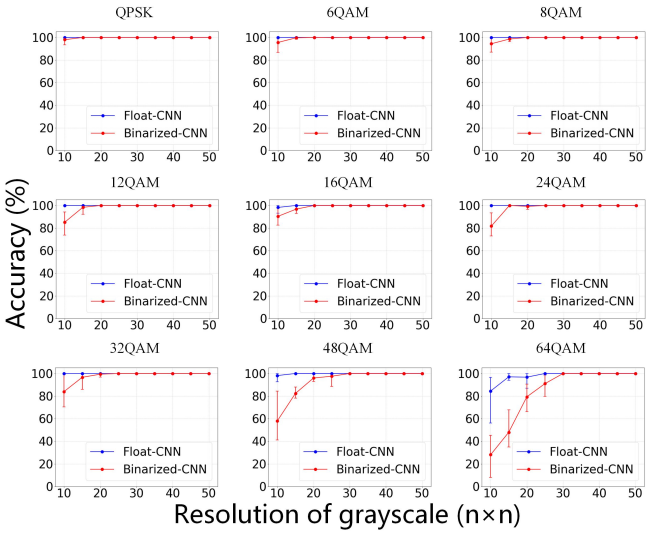

Fig. 6. Accuracy of OSNR estimation by B-CNN and F-CNN vs. resolutions of grayscale (average, maximum and minimum value from five random initialization).

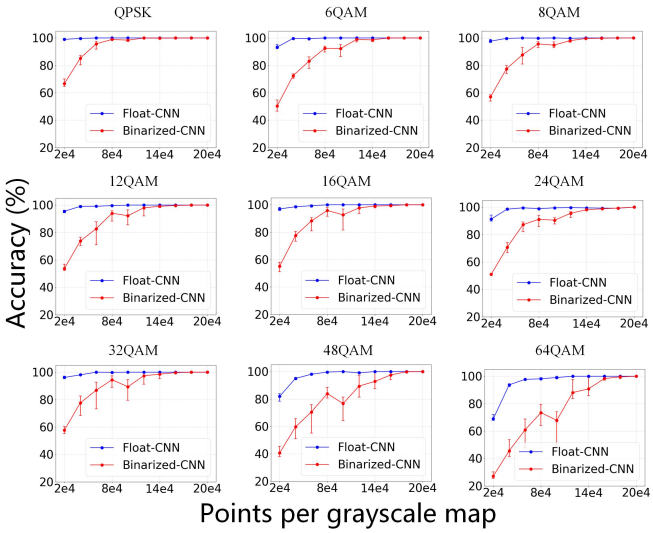

Fig. 7. Accuracy of OSNR estimation by B-CNN and F-CNN vs. points per grayscale map (average, maximum and minimum value from five random initialization).

a register (SWAR) [19]. The basic idea of SWAR is to concatenate groups of 32 binary variables into 32-bit registers, and thus obtain a 32-times speed up on bitwise operations (e.g. XNOR). In addition, hardware dedicated to BNNs is also under investigation [20], [21], which would further improve the performance of BNNs. Therefore, we believe that the proposed technique using B-CNN will be comparable to those of existing joint MFI and OSNR monitoring methods.

C. *Robustness of B-CNN*

1) Polarization multiplexing: To investigate the robustness of proposed method, we further evaluate the proposed method in DP-QAM system with different transmission distances. Same as subsection *A*, we first set the transmission distance at

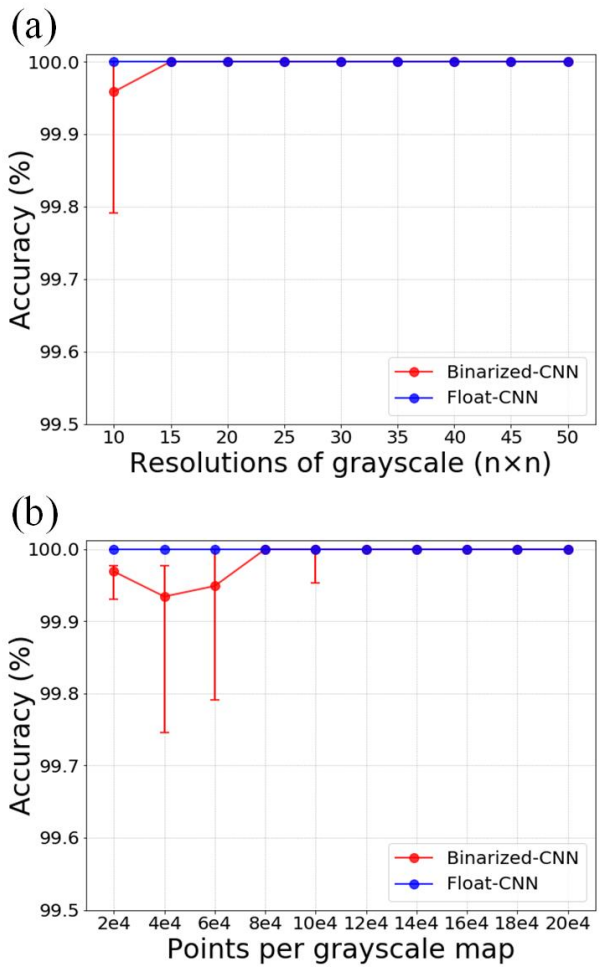

Fig. 8. (a) Accuracy of MFI vs. resolutions of grayscale, (b) Accuracy of MFI vs. points per grayscale map (average, maximum and minimum value from five random initialization).

240 km and the resolution of maps at 40×40 and each map contains 2×10^5 signal points. In this case, the accuracies of MFI and OSNR estimation are all 100%. Therefore, the proposed method is not affected by the way of polarization multiplexing.

2) Transmission distance: We change the transmission distance to 160 km and 80 km and evaluated the performance of B-CNN. We set the resolution of maps at 40×40 and each map contains 2×10^5 signal points. The accuracies of MFI and OSNR estimation remain 100%. Since the above results are obtained by training B-CNN at different transmission distance respectively, we further explore the robustness of B-CNN to transmission distances by training it only once using all data from three different transmission distances. The accuracy of MFI remains 100%, which indicates that B-CNN is robust to transmission distance in MFI. The accuracies of OSNR estimation slightly decline as shown in Fig. 9. However, this result may be further improved by using more data from more different transmission distances.

3) Fiber nonlinearity: We change the input optical power to the fiber to 3 dB, 5 dB and 7 dB with transmission of 240 km. As shown in Fig. 10, due to the fiber nonlinearity, the

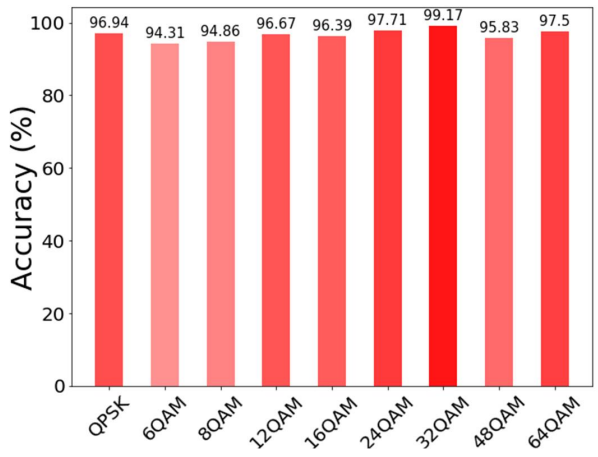

Fig. 9. Accuracy of OSNR estimation by B-CNN trained on data from three different distances (80 km, 160 km, and 240 km).

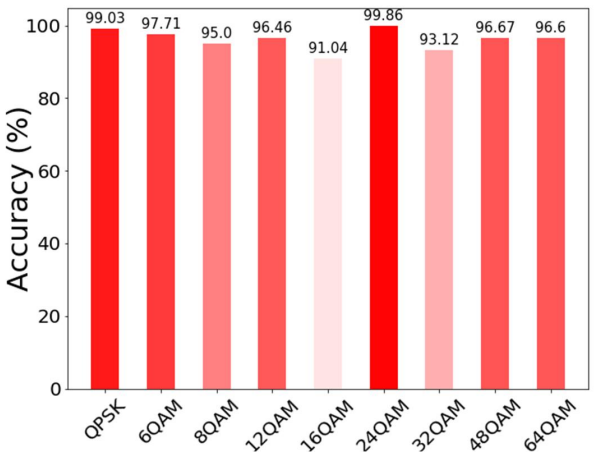

Fig. 11. Accuracy of OSNR estimation by B-CNN trained on data from three different launched powers (3 dB, 5 dB, and 7 dB).

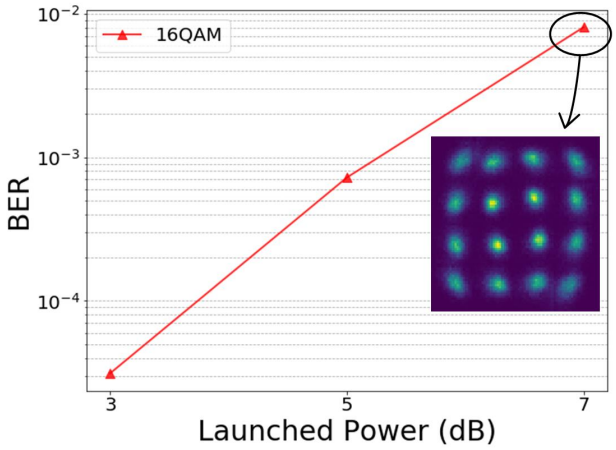

Fig. 10. BER vs. Launched Power for 16QAM.

bit error rate (BER) drops as the launched power increase and there exist distortions in the constellation maps. Same as the previous part, we set the resolutions at 40×40 and each map contains 2×10^5 signal points. We train B-CNN respectively using data obtained from the DP-QAM system with launched power at 3 dB, 5 dB, and 7dB. The accuracies of MFI and OSNR estimation are all 100%. For the same consideration as the previous part, we mix the data from three different launched powers and train the B-CNN once. The accuracy of MFI remains 100% and the accuracies of OSNR estimation are also reduced as shown in Fig. 11.

4) Discussion: According to the results above, we can conclude that B-CNN is able to achieve high performance if we retrain it every time we change the configuration of the system. However, the performance will be reduced if we mix data from the system with different parameters. This phenomenon leaves us to consider how to construct a big data set with features from different systems to train the model and improve its robustness, since the cost will be great if we retrain the model every time the parameters of the system are changed.

V. Experimental Setup and Results

The experimental setup for the demonstration of the proposed OSNR monitoring and MFI technique is shown in Fig. 12. We generate 12.5 Gbaud flexible-QAM (QPSK, 6QAM, 8QAM, 12QAM, 16QAM, 24QAM, 32QAM, 48QAM and 64QAM) optical signals by modulating a carrier signal, provided by an external cavity laser (ECL), using I/Q modulators which are driven by multi-level electrical signals. The center wavelength of ECL is 1552.52 nm and its linewidth is 100 kHz. Polarization multiplexing is then realized by utilizing polarization beam splitters (PBSs), polarization beam combiners (PBCs), and optical delay lines. The resulting signals are amplified using an erbium-doped fiber amplifier (EDFA) and sent over a 5 km long standard single-mode fiber (SSMF). A variable optical attenuator (VOA) is utilized to alter OSNRs of flexible-QAM signals in the range of $10 \sim 25$ dB for QPSK, 6QAM, 8QAM, 12QAM, $15 \sim 30$ dB for 16QAM, 24QAM, and $20 \sim 35$ dB for 32QAM, 48QAM, 64QAM. The optical signals at the output of EDFA are filtered using a 0.6 nm optical band-pass filter (OBPF) and then detected by a coherent receiver. The electrical signals after optical-to-electronic (O/E) conversion are sampled by utilizing an oscilloscope with 50 Gsamples/s sampling rate and 2×10^6 samples are collected which are then processed offline using a DSP core. As clear from Fig. 8, the proposed technique processes polarization de-multiplexed signals after CMA equalization. We generate a large data set encompassing 14400 constellation diagrams corresponding to different modulation formats and OSNRs. This data set are divided into training and testing data sets by randomly selecting 70% (i.e. 10080) and 30% (i.e. 4320) of overall constellation diagrams, respectively. Finally, these constellation images are transformed to grayscale maps and input to the B-CNN for joint MFI and OSNR monitoring.

Table II-III summarizes the MFI and OSNR estimation results of experimental data. As expected, the MFI accuracy

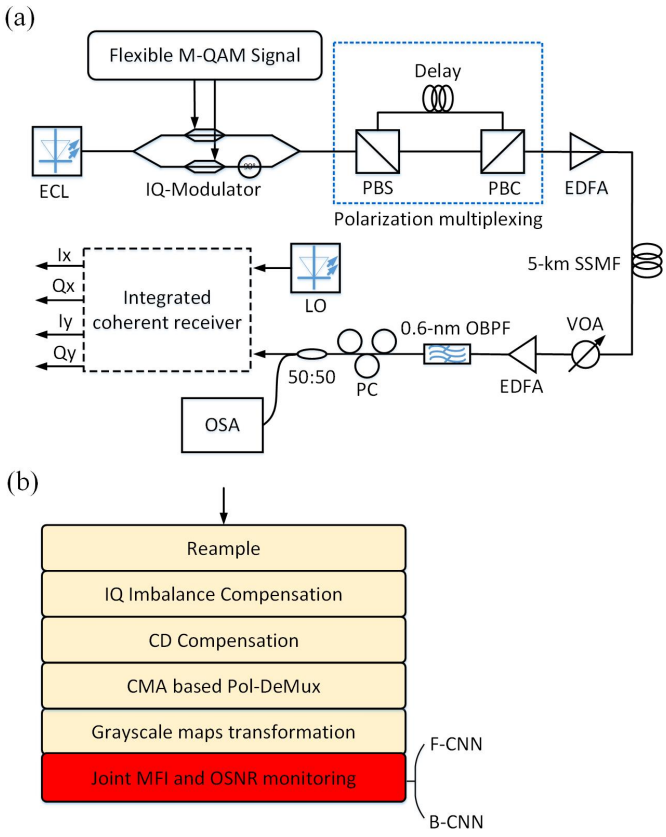

Fig. 12. Experimental setup for joint MFI and OSNR monitoring. ECL: external cavity laser; PBS: polarization beam splitter; PBC: polarization beam combiner; VOA: variable optical attenuator; OBPF: optical band-pass filter; PC: polarization controller; OSA: optical spectrum analyzer.

is 100%, and the accuracy of OSNR estimation varies from 95.83% to 100% among nine modulation formats. For 64QAM signals, the accuracy of OSNR estimation is lowest (i.e. 95.83%), while the accuracy of all other modulation formats are higher than 99.38%. Fig. 13. shows the estimated OSNR by B-CNN as a function of measured OSNR, the red circles denote the averaged value of estimated OSNR over test data set. It is obvious that the large estimation errors mainly arise in 64QAM signals, the maximum error of all other modulation formats are below 0.5 dB, while the maximum error of 64QAM is 1.14 dB. The results of the experimental data prove that the proposed scheme can be successfully utilized to identify modulation format of flexible-QAM signals and estimate OSNR of various modulation formats simultaneously.

VI. Conclusion

In this paper, we have proposed and experimentally verified a joint OSNR monitoring and MFI technique in digital coherent receivers by processing ring constellations grayscale maps based on B-CNN. When the resolution of grayscale maps is set at 40×40 and each map contains 2×10^5 signal points. the simulated and experimental results show that the MFI accuracy of nine QAM signals in coherent system is 100%. In experiment, the accuracy of OSNR estimation of 64QAM is 95.83% and the accuracies of other eight modulation formats are higher than 99.38%. Compared with F-CNN, B-CNN has less than 1/3 memory consumption and requires 1/2 epochs to get stable performance. The improvement can be more noticeable with dedicated hardware. The results also show that B-CNN has lower tolerance of resolution and points number of grayscale maps compared with F-CNN. Nevertheless, this proposed technique is attractive for real-time OPM in future optical network because of its low memory consumption and high computation efficiency.

References

[1] 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," *Proceedings of the IEEE*, vol. 102, no. 9, pp. 1317–1337, 2014.

[2] H. Khodakarami, B. S. G. Pillai, B. Sedighi, and W. Shieh, "Flexible optical networks: An energy efficiency perspective," *Journal of Lightwave Technology*, vol. 32, no. 21, pp. 3356–3367, 2014.

[3] R. Borkowski, R. J. Durán, C. Kachris, D. Siracusa, A. Caballero, N. Fernández, D. Klonidis, A. Francescon, T. Jiménez, J. C. Aguado *et al.*, "Cognitive optical network testbed: Eu project chron," *Journal of Optical Communications and Networking*, vol. 7, no. 2, pp. A344–A355, 2015.

[4] J. Zhang, W. Chen, M. Gao, Y. Ma, Y. Zhao, and G. Shen, "Intelligent adaptive coherent optical receiver based on convolutional neural network and clustering algorithm," *Optics express*, vol. 26, no. 14, pp. 18 684–18 698, 2018.

[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," *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 525–543, 2015.

[6] M. A. Jarajreh, E. Giacoumidis, I. Aldaya, S. T. Le, A. Tsokanos, Z. Ghassemlooy, and N. J. Doran, "Artificial neural network nonlinear equalizer for coherent optical ofdm," *IEEE Photonics Technology Letters*, vol. 27, no. 4, pp. 387–390, 2014.

[7] J. Thrane, J. Wass, M. Piels, J. C. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected pdm-qam signals," *Journal of Lightwave Technology*, vol. 35, no. 4, pp. 868–875, 2016.

[8] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bülow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end deep learning of optical fiber communications," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, 2018.

[9] 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," *Optics express*, vol. 25, no. 15, pp. 17 767–17 776, 2017.

[10] 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," *Optics express*, vol. 25, no. 15, pp. 17 150–17 166, 2017.

[11] 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 Journal*, vol. 10, no. 5, pp. 1–12, 2018.

[12] 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," *Optics express*, vol. 27, no. 8, pp. 11 281–11 291, 2019.

[13] Y. Cheng, S. Fu, M. Tang, and D. Liu, "Multi-task deep neural network (mt-dnn) enabled optical performance monitoring from directly detected pdm-qam signals," *Optics Express*, vol. 27, no. 13, pp. 19 062–19 074, 2019.

[14] 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," *Optics Express*, vol. 27, no. 7, pp. 9403–9419, 2019.

[15] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," *arXiv preprint arXiv:1710.09282*, 2017.

[16] M. Courbariaux, I. Hubara, D. Soudry, R. El-Yaniv, and Y. Bengio, "Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1," *arXiv preprint arXiv:1602.02830*, 2016.

[17] A. Gulli and S. Pal, *Deep Learning with Keras*. Packt Publishing Ltd, 2017.

TABLE II
MFI RESULTS OF EXPERIMENTAL DATA

		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%

TABLE III
OSNR ESTIMATION RESULTS OF EXPERIMENTAL DATA

Modulation format	QPSK	6QAM	8QAM	12QAM	16QAM	24QAM	32QAM	48QAM	64QAM
Accuracy	100%	99.79%	100%	99.58%	100%	99.38%	99.79%	99.58%	95.83%

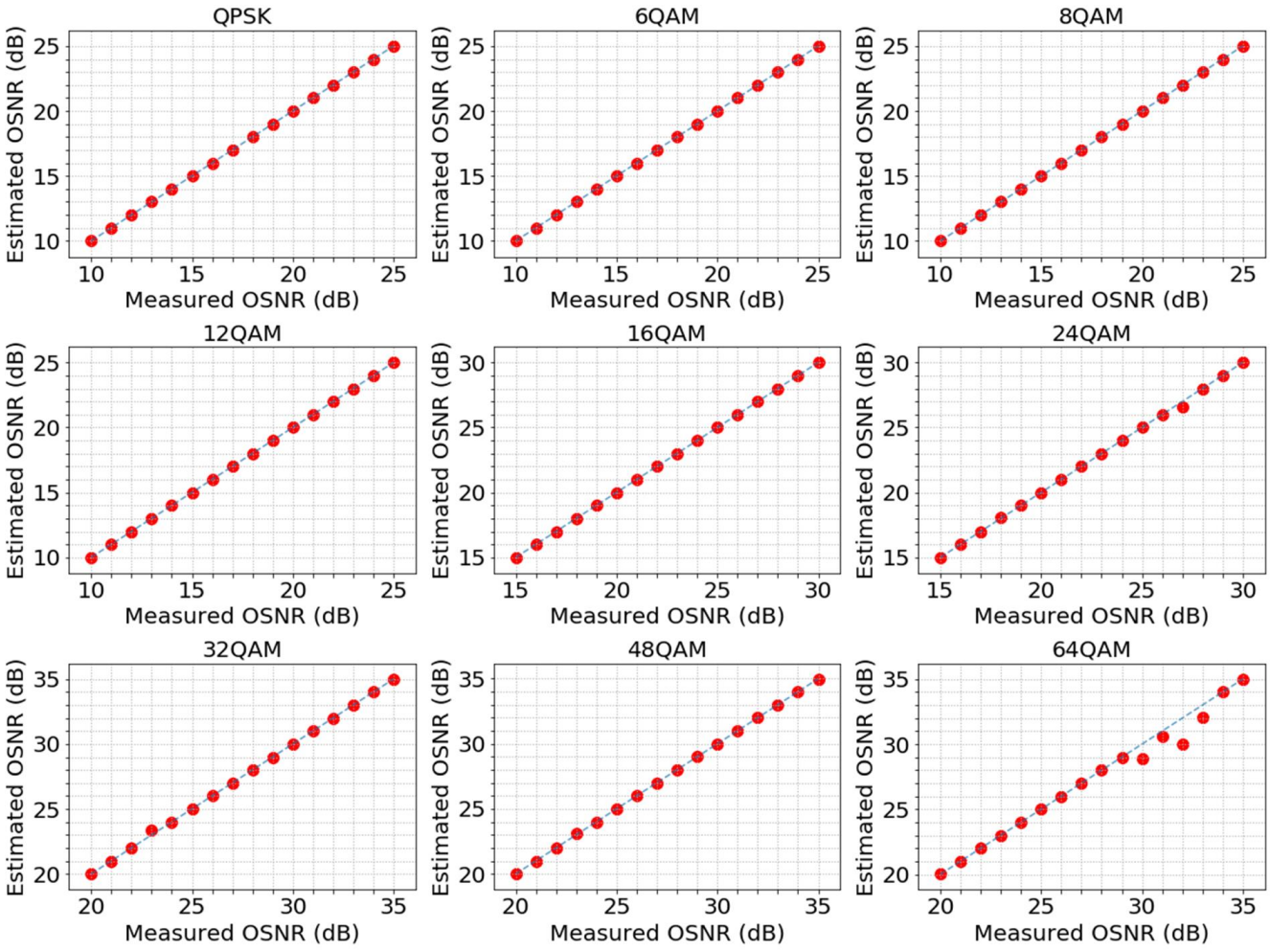

Fig. 13. Experimental results for OSNR estimation for different modulation formats by B-CNN.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
[20] R. Zhao, W. Song, W. Zhang, T. Xing, J.-H. Lin, M. Srivastava, R. Gupta, and Z. Zhang, "Accelerating binarized convolutional neural networks with software-programmable fpgas," in *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*. ACM, 2017, pp. 15–24.
[21] K. Ando, K. Ueyoshi, K. Orimo, H. Yonekawa, S. Sato, H. Nakahara, S. Takamaeda-Yamazaki, M. Ikebe, T. Asai, T. Kuroda *et al.*, "Brein memory: A single-chip binary/ternary reconfigurable in-memory deep neural network accelerator achieving 1.4 tops at 0.6 w," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 4, pp. 983–994, 2017.