Low complexity Intelligent OPM using binarized neural networks

Yilun Zhao[1], Zhenming Yu[1,*], Zhiquan Wan[1], Shaohua Hu[2], Liang Shu[1], Jing Zhang[2] and Kun Xu[1]

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

2 Key Laboratory of Optical Fiber Sensing and Communications, University of Electronic Science and Technology of China, Chengdu, China

**Author e-mail address: yuzhenming@bupt.edu.cn*

Abstract: We propose and experimentally demonstrate a novel optical performance monitoring (OPM) method using a binarized convolutional neural network (B-CNN). It reduces memory consumption and training overhead significantly without performance loss.

1. Introduction

With the increasing of the demand for various data services, optical network is evolving from current fixed architecture to future flexible one. In order to reduce operating costs, ensure optimum resources utilization and guarantee adequate operation and management of flexible optical networks, it is essential to continuously monitor various network performance parameters, which is referred to as optical performance monitoring (OPM) [1,2].

Among all the parameters of physical layer, optical signal-to-noise ratio (OSNR) and modulation format type are especially vital for coherent links [3]. Recently, the combination of OSNR estimation and modulation format identification (MFI) is becoming a trend of development. Many researches have employed machine learning (ML) techniques to achieve joint OSNR monitoring and MFI [3-7]. In [4], D. Wang *et al*. proposed an intelligent constellation diagram analyzer using convolutional neural network (CNN) and proved that CNN outperformed other ML techniques such as decision tree, multilayer perceptron (MLP), *etc*. In addition, D. Wang *et al*. further proposed a CNN-based OPM at intermediate node for cost efficiency [5]. Although CNN has a strong ability in joint OSNR monitoring and MFI, existing schemes are computationally expensive and memory intensive, which is not proper for the fulture receivers with low memory resources and strict latency requirements. In this paper, we propose a novel OPM technique using binarized convolutional neural network (B-CNN). The B-CNN is constructed based on the binarized neural networks (BNNs) [8] with binary weights and activations at run-time during the forward pass, which drastically reduces memory size and is expected to substantially improve power efficiency. The performance of the proposed technique is verified in experiment.

2. Principle

When training a BNN [8], both the weights and the activations are constrained to either +1 or -1. These two values are very advantageous from a hardware perspective. We select a deterministic function to transform the real-valued variables into these two values: $x^b = \mathrm{Sign}(x) = (+1 \text{ if } x \geq 0, -1 \text{ otherwise})$, 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. In order to address this problem, we propagate the gradient through *hard tanh*, which is the following piece-wise linear activation function: $\mathrm{Htanh}(x) = \mathrm{Clip}(x, -1, 1) = \max(-1, \min(1, x))$. For hidden units, we use the *sigh* 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.

3. Details of B-CNN model

In our scheme, the B-CNN consists of two convolutional layers, one maxpooling layer and two fully-connected layers (as shown in Fig. 1). For comparison, the float valued convolutional neural network (F-CNN) with the similar architecture is adopted. In F-CNN, the filter numbers for the two convolutional layers are 32 and 64 respectively and 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.

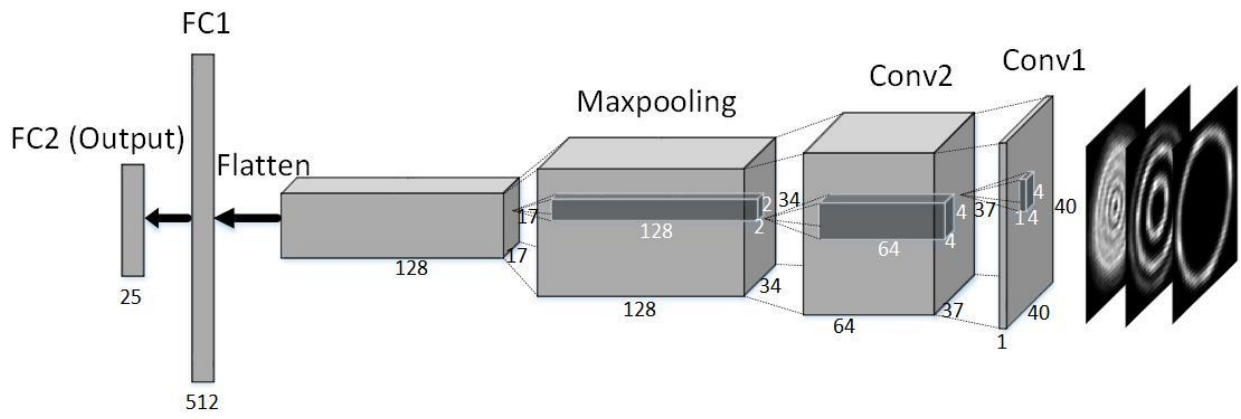

Fig. 1. Architecture of B-CNN. Conv: convolutional layer; FC: fully-connected layer.

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 $-\sum_{c=1}^{M} y_{o,c} \log\left(p_{o,c}\right)$, 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 so that the *cross-entropy* fail to work. Therefore, we choose *squared-hinge* function as the loss function: $L(y,t) = \max\left(0,(1-yt)^2\right)$, 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.

4. Experimental Setup and Results

4.1. Experimental setup and data preparation

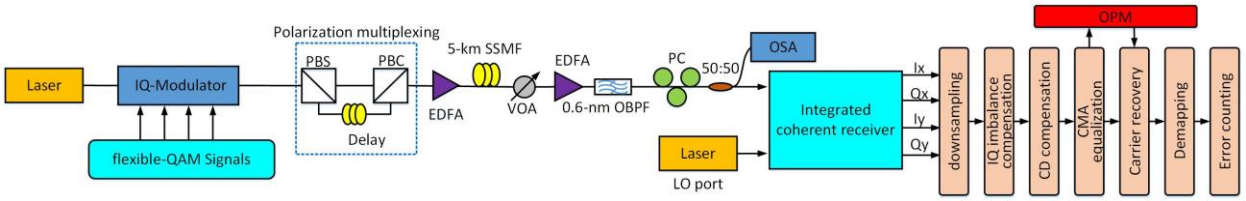

Fig. 2. Experimental setup.

The proposed OPM method is verified in experiment, as show in Fig. 2. At the transmitter side, we generate 12.5 Gbaud dual-polarization-flexible-QAM 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 line width 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. B-CNN-based OPM is integrated into the digital signal processing (DSP) flow. After downsampling, IQ imbalance compensation, chromatic dispersion (CD) compensation and polarization demultiplexing using constant modulus algorithm (CMA), the received signal is utilized to generate the ring grayscale constellation maps for B-CNN to recognize modulation formats and estimate OSNR values.

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 (10~25 dB for QPSK, 6QAM, 8QAM, 12QAM, 15~30 dB for 16QAM, 24QAM and 20~35 dB for 32QAM, 48QAM, 64QAM). In the data set, each map has a label vector, we set the first 9 bits to denote the modulation formats and the other 16 bits to denote the 16

OSNR values so that the total number of bits is 25. Table 1 illustrates the relationship between modulation formats and corresponding label vectors by listing the labels of QPSK, 16QAM, and 64QAM signals.

Table 1. Examples of label vectors

Signal Type	QPSK, 15 dB	16QAM, 20 dB	64QAM, 35 dB
First 9 bits	[+1,-1,-1,-1,-1,-1,-1,-1,-1]	[-1,-1,-1,-1,+1,-1,-1,-1,-1]	[-1,-1,-1,-1,-1,-1,-1,-1,+1]
Last 16 bits	[+1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1]	[-1,-1,-1,-1,-1,+1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1]	[-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,+1]

4.2. Results analysis

The experimental results are shown in Fig. 3. We compare the performance of the proposed OPM using B-CNN and the OPM using F-CNN. For both OPMs using B-CNN and F-CNN, the accuracy of MFI is 100%. As for OSNR estimation, both methods can achieve accuracies higher than 95% for nine modulation formats. As shown in Fig. 3(a), for QPSK, 6QAM, 8QAM, 12QAM, 16QAM, 24QAM, 32QAM and 48QAM, the OPM using B-CNN can achieve similar accuracy of OSNR estimation compared with the OPM using F-CNN. As for 64QAM, the accuracy of OSNR estimation using B-CNN is 95.83 while the accuracy of OSNR estimation using F-CNN is 98.54. As shown in Fig 3(b), the parameters size of B-CNN is 3.14 MB and the parameters size of F-CNN is 10.23 MB. We also measure the accuracies of OSNR estimation versus number of epochs for B-CNN and F-CNN simultaneously. We choose the accuracies of OSNR estimation for 64QAM to depict the accuracy as a function of number of epochs. Fig. 3(c) shows that B-CNN requires 9 epochs to achieve stable performance and F-CNN requires 20 epochs. As a conclusion, B-CNN can achieve similar performance compared with F-CNN, while the memory consumption is reduced by 2/3 and the training overhead is reduced by 1/2.

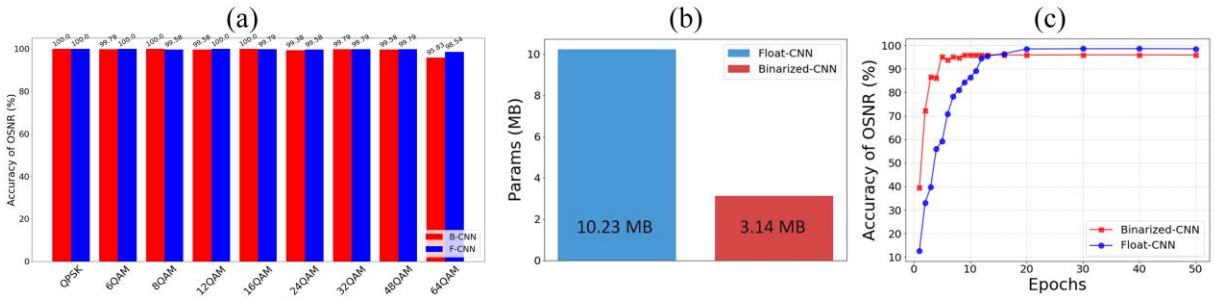

Fig. 3. Experimental results. (a) The accuracies of OSNR estimation using B-CNN and F-CNN; (b) Comparison of F-CNN and B-CNN in parameters size; (c) Accuracy of OSNR estimation in epochs number using B-CNN and F-CNN for 64QAM.

5. Conclusion

In this paper, we propose a low-complexity OPM method by processing ring constellation grayscale maps using B-CNN for DP-flexible-QAM coherent system. The proposed method is experimentally verified with little performance penalty compared with the OPM using F-CNN, while the memory consumption is reduced by ~2/3 and the training overhead is reduced by ~1/2.

6. Acknowledgements

This work is supported by the NSFC Program (No. 61431003, 61901045, 61625104, 61601049, 61821001). The work is also part of the National Key R&D Program of China (No. 2018YFB2201803); Fundamental Research Funds for the Central Universities.

7. References

[1] I. Tomkos, et al., Proceeding IEEE, 104(9), 2014.
[2] Z. Dong, et al., IEEE J Lightwave Technol, 34(2), 2016.
[3] F. N. Khan, et al., Optic Express, 25(15), 2017.
[4] D. Wang, et al., Optic Express, 25(15), 2017.
[5] D. Wang, et al., Optic Express, 27(7), 2019.
[6] Z. Wan, et al., Optic Express, 27(8), 2019.
[7] Y. Cheng, et al., Optic Express, 27(13), 2019.
[8] M. Courbariaux, et al., arXiv, 2016.