IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 28, 2019, accepted August 27, 2019, date of publication September 2, 2019, date of current version September 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939043

Feature Fusion-Based Multi-Task ConvNet for Simultaneous Optical Performance Monitoring and Bit-Rate/Modulation Format Identification

XIAOJIE FAN[1], LINA WANG[1], FANG REN[1], YULAI XIE[2], XIANG LU[3], YIYING ZHANG[1], TIANWEN ZHANGSUN[1], WEI CHEN[1], AND JIANPING WANG[1]

[1]School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
[2]Hitachi (China) Research and Development Company Ltd., Beijing 100083, China
[3]Information Center, Guizhou Power Grid Company Ltd., Guiyang 550002, China

Corresponding author: Lina Wang (wln_ustb@126.com)

This work was supported in part by the Fundamental Research Funds for Central University, China, under Grant FRF-TP-19-016A2, in part by the National Natural Science Foundation of China (NSFC) under Grant 61671055 and Grant 61605004, and in part by the State Key Laboratory of Advanced Optical Communication Systems Networks, China.

ABSTRACT We propose a novel feature fusion based multi-task convolutional neural network (ConvNet) for simultaneous bit-rate and modulation format identification (BR-MFI) and optical performance monitoring (OPM) in heterogeneous fiber-optic networks. The proposed multi-task ConvNet fuses the intermediate layers through the convolutional operation and then trains multi-task losses on the fused feature. In addition to traditional multi-task ConvNet's ability of the feature extraction and sharing, our multi-task ConvNet is able to capture both global and local information of phase portraits and has good performance on OPM and BR-MFI tasks in a short processing time (∼51 ms). The simulation results of six signals (consisted of two bit-rates and three modulation formats) demonstrate the root-mean-square (RMS) errors of the optical signal-to-noise ratio (OSNR), chromatic dispersion (CD) and differential group delay (DGD) are 0.81 dB, 1.52 ps/nm and 0.32 ps, respectively. Meanwhile, the 100% classification accuracy can be obtained for BR-MFI. Besides, the effects of the fused feature shape, the location of feature extracted for fusion, the transmitter variations and fiber nonlinearity on the performance of the proposed technique are thoroughly investigated.

INDEX TERMS Bit-rate and modulation format identification (BR-MFI), optical performance monitoring (OPM), convolutional neural network (ConvNet), feature fusion.

I. INTRODUCTION

OPM as well as BR-MFI is planned to be an important part of the future optical networks for the purpose of monitoring the quality of optical signals precisely. The future dynamic optical networks are designed to transmit various signal rates and modulation formats to meet the needs of signal transmission, and the signals may accumulate different amounts of impairments during the transmission [1], [2]. Many proposed OPM techniques presume the prior knowledge of the signal's format and bit-rate or attain these information from the upper-layer protocols to monitor the network impairments [3]. Nevertheless, it's unpractical to add extra cross-layer communication for the OPM devices since they have limited tolerance of complexity. Hence, it's meaningful to develop OPM techniques which can monitor various impairments under different formats and bit-rates without any prior information. Additionally, BR-MFI is also important since the OPM techniques may be signal type dependent. And the identified signal type information is helpful for the digital coherent receivers to choose a proper carrier recovery module. Therefore, in order to monitor critical optical performance parameters and identify modulation formats/signal rates in a real-time way, it is significant to develop OPM and BR-MFI techniques for the optical network.

Over the past few years, various machine learning algorithms have been applied for OPM or BR-MFI. Previous works used various algorithms from the community of machine learning, including the support vector machine (SVM) [4], the k-nearest neighbors (KNN) [5], as well as the back-propagation artificial neural network (BP-ANN) [3], [6]. However, these works were proposed only for OPM

The associate editor coordinating the review of this article and approving it for publication was Tianhua Xu.

 This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

or BR-MFI alone, since they regard the OPM and BR-MFI as separate tasks and ignore the potential relevance between them, which is inefficient. Moreover, many machine learning algorithms such as principal component analysis (PCA) [3], [7], BP-ANN [8] were proposed for the joint optimization of OPM and BR-MFI. The problem of joint optimization of OPM and BR-MFI can be divided into the impairments monitoring task, the modulation format classification task and the bit-rate classification task. However, these works mainly focus on OSNR parameter, regardless of other critical optical performance parameters. Furthermore, all of the machine learning algorithms used by these works lack the capability to extract and share features. Specifically, rich domain expertise is needed to design a feature extractor since machine learning algorithms cannot process raw data directly, and the extracted features cannot share helpful information among each task. Therefore, it is necessary to develop more advanced algorithms to avoid the drawbacks of machine learning algorithms.

Lately, deep learning (DL) with the strong capability of automatic feature extraction has attracted more and more attention [9]–[11]. Some joint OSNR estimation and format classification methods based on ConvNet and eye diagrams have been proposed [12], [13]. However, the above methods based on conventional ConvNet only solved the problem of feature extraction except feature sharing. We believe that OPM and BR-MFI are not a standalone problem, but their task performance can be influenced by each other. For example, different types of signals have different tolerance to the specific impairments. Efficiently extracting and sharing the features provided by the relevant tasks like BR-MFI and OPM can boost the performance of the individual tasks. Guided by this idea, our previous work proposed a ConvNet-based multi-task learning method to attain the capability of feature sharing for the sake of joint optimization of OPM and BR-MFI [14]. Thanks to the multi-task learning method, our multi-task ConvNet attained both the capability of feature sharing and feature extraction [15], [16].

However, our previous work still can be improved in the following aspects. (1) Reduce the cost of the devices. Our previous work used asynchronous delay-tap sampling (ADTS) method which requires a couple of high-bandwidth samplers to obtain phase portraits [17]. Instead, we can use asynchronous single channel sampling (ASCS) method which only require single-tap sampling to obtain phase portraits [18], [19]. (2) Use feature fusion to improve performance. The fact which is overlooked is that the distribution of information on features is hierarchical throughout the ConvNet as proved in [20]. Lower layers respond to edges, corners, and the density of sample points presented in the initial phase portrait, and hence contains better impairment features. They would be more suitable for learning impairments monitoring task since optical impairments would directly affect the shape and amplitude of the signal and change the distribution of sample points in phase portraits. On the other hand, higher layers carry increasingly less information

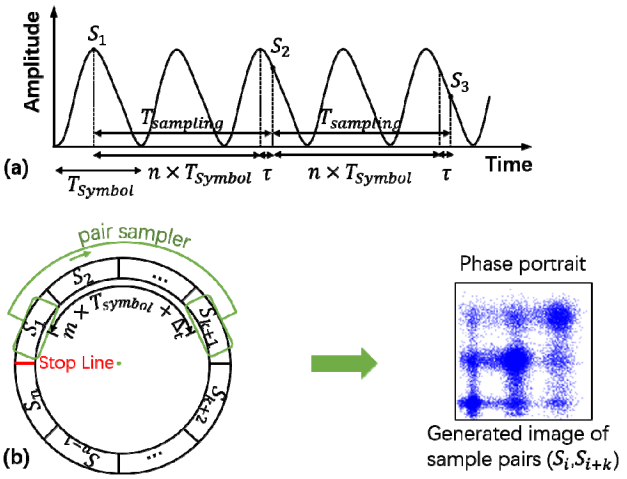


FIGURE 1. (a) Schematic of the ASCS. (b) Generation of a phase portrait using sample pairs from the sample sequence.

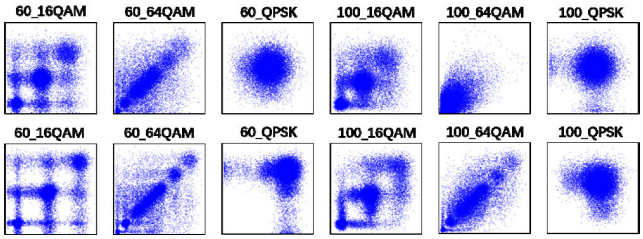


FIGURE 2. Phase portraits obtained by ASCS for all types of signals (three formats and two bit-rates) in the scenario of various impairments. The first row corresponds to OSNR = 10 dB and without DGD and CD. The second row corresponds to OSNR = 22 dB, DGD = 8 ps, and CD = 50 ps/nm.

about the visual contents of the phase portrait, and increasing more information related to the class of the phase portrait. Therefore, they would be more suitable for learning abstract concept related classification tasks which are desired for modulation format classification and bit-rate classification. Obviously, in our previous work, it is imperfect to train a multi-task ConvNet for joint OPM and BR-MFI using features only from the last layer. Therefore, we extract and transform features from different layers of multi-task ConvNet to a common subspace. Then, based on the fused feature, we train three tasks (the impairments monitoring task, the modulation format classification task, and the bit-rate classification task) simultaneously based on the overall loss function. Both the global and local information of phase portrait is captured by our method through this way.

In this paper, we improve our previous work and propose a novel feature fusion based multi-task ConvNet which uses fused features for simultaneous OPM and BR-MFI tasks in conjunction with phase portraits obtained through ASCS. The use of ASCS reduces the cost of the monitoring devices. Our novel multi-task ConvNet improves task performance. Numerical simulations are performed for 60/100 Gbps quadrature phase-shift keying (QPSK), 60/100 Gbps 16 quadrature amplitude modulation (16QAM), and 60/100 Gbps 64QAM signals in the scenario of various

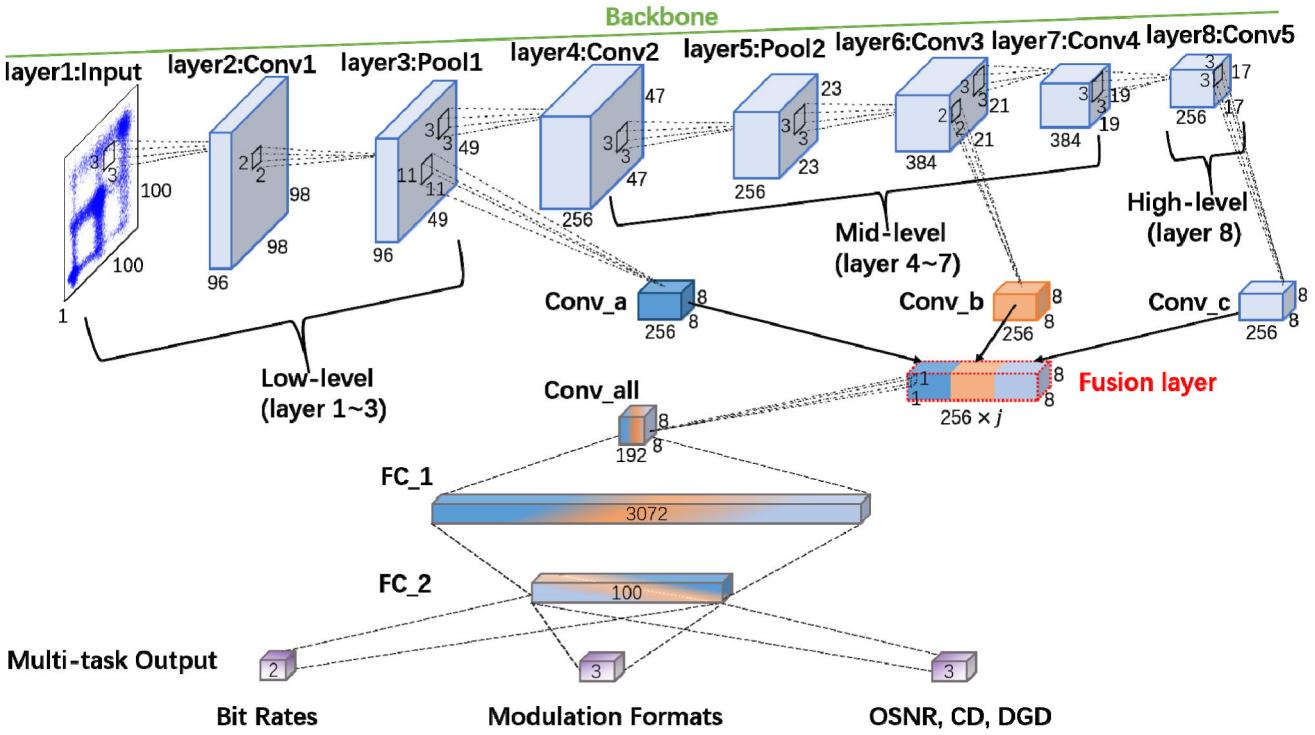


FIGURE 3. The architecture of the feature fusion based multi-task ConvNet. The "Backbone" part is consisted of 8 consecutive layers. These layers can be divided into three levels: The low-level (layer 1-3), the mid-level (layer 4-7) and the high-level (layer 8). Features ("Conv_a", "Conv_b" and "Conv_c") extracted from different levels are concatenated in the fusion layer. The parameter $j \in \{1, 2, 3\}$ in the fusion layer represents the number of fused features. There is no feature to fuse and only exist the "Conv_c" feature when j equals 1. The "Conv_c" is concatenated with one of the "Conv_a" and the "Conv_b" when j equals 2. All three feature are concatenated together when j equals 3. Finally, the output of the fusion layer is send to the fully connected layers to predict the results. The shape parameter of each layer is shown in the figure.

impairments such as OSNR, DGD, and CD, validate simultaneous OPM and BR-MFI with good accuracies.

II. METHODS

A. ASYNCHRONOUS SINGLE CHANNEL SAMPLING

The schematic diagram of ASCS is shown in Fig. 1(a). After the direct detection of optical signal by the photodetector (PD), the transformed electrical signal amplitude is sampled asynchronously with a slow rate. The time interval T_{sampling} between adjacent samples can be expressed as

$$T_{sampling} = nT_{symbol} + \tau \tag{1}$$

where n is an integer, T_{symbol} is the symbol period and is much bigger than the time interval τ. Furthermore, between samples S_i and S_{i+k}, the total time interval T_{total} can be expressed as

$$T_{total} = knT_{symbol} + k\tau = mT_{symbol} + \Delta_t \tag{2}$$

where Δ_t is the cumulative remaining time interval of τ while m is an integer. Consider $S_1, S_2, \cdots S_N$ be a discrete sample sequence acquired by ASCS, we connect the sample sequence end to end, then rotate clockwise the "pair sampler" to attain sample pair (S_i, S_{i+k}) started from the head of the sequence, as shown in Fig. 1(b). It is noted that the "pair sampler" with fixed sample spacing cannot surpass the stop line.

Next, the sample pairs $(S_1, S_{1+k}), (S_2, S_{2+k}), \cdots (S_{N-k}, S_N)$ are used to generate the phase portrait. The ADTS needs a pair of high-bandwidth samplers and a physical delay line, but the ASCS only need one sampler, which is very economical. The phase portraits of all types of signals under various transmission impairments are shown in Fig. 2. It is evident that the signal types as well as optical impairments have a great influence on patterns of phase portraits such as shape, edge, and the density of points. Hence, we can design novel multi-task ConvNet which makes full use of the global and local image information for simultaneous OPM and BR-MFI tasks with higher accuracies.

B. FEATURE FUSION BASED MULTI-TASK CONVNET

Two theories are used to create our network: (1) The distribution of the features is hierarchical in ConvNet. The features of the lower layer are informative for impairments monitoring task, and the features of the higher layer are more appropriate for the abstract classification task. (2) Multi-task learning can improve the performance of the individual task by the feature sharing, which has been confirmed by our previous work [14]. Therefore, for simultaneous OPM and BR-MFI, it is meaningful to fuse the intermediate layers' features and learn all tasks based on the fused feature.

The structure of our network is illustrated in Fig. 3. The "Backbone" part of the network is consisted of eight layers started from the first "Input" layer to the eighth "Conv5" layer. The input phase portrait, which is a 100×100

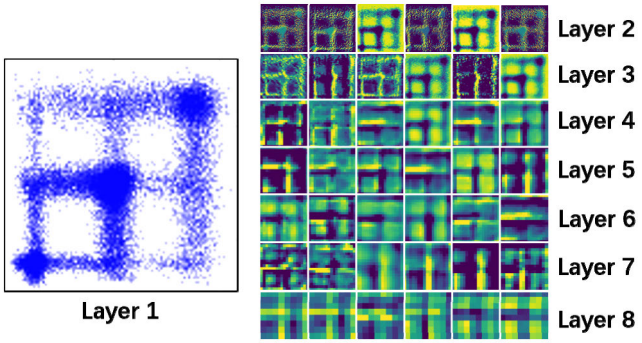


FIGURE 4. Several feature maps of each layer in the "Backbone" part.

gray-scale image, would be processed successively by the layers in the "Backbone" part, and the features of different layers would be produced. Then, we extract features from the "Pool1", "Conv3", and "Conv5" layers. Due to the different dimension of feature maps from these layers ("Pool1": $49 \times 49 \times 96$, "Conv3": $21 \times 21 \times 384$, "Conv5": $17 \times 17 \times 256$), we add "Conv_a", "Conv_b", and "Conv_c" convolutional layers to "Pool1", "Conv3", and "Conv5" layers respectively, to attain feature maps with the same dimension of $8 \times 8 \times 256$. Then, we use the fusion layer to concatenate the features to form a feature map with the dimension of $8 \times 8 \times (256 \times j)$, where the parameter j is the number of feature maps to be concatenated in the fusion layer. In order to fuse the concatenated features, a convolution layer named "Conv_all" with the 1×1 kernel is added to attain feature maps with the dimension of $8 \times 8 \times 192$. The fully connected layer "FC_1" is connected to "Conv_all" layer. The "FC_2" layer with the dimension of 100 is fully connected to "FC_1" layer. Finally, three fully connected layers with the dimension of 2, 3, and 3 are parallel added to the "FC_1" layer to predict the bit-rate classification task, the modulation format classification task and the impairments monitoring task, respectively.

We present several feature maps of each layer in "Backbone" part, as shown in Fig. 4. It is clear that the feature maps are distributed hierarchically from the first to eighth layers. We can roughly divide these feature maps into three levels. The first level is the low-level features (1-3 layers) containing lots of visual information about edges, corners, and the density of points, and the second level is the mid-level features (4-7 layers) becoming increasingly abstract and less visually interpretable. This means the mid-level starts to encode concepts such as "format" or "bit-rate". The third level is the high-level features (8th layer) where no visual information exists but abstract concepts. The three levels are also marked in Fig. 3.

Given total N training samples, the training data of the input phase portraits can be denoted as $\{\mathrm{x}_i\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^{100\times100\times1}$. The labels of the impairments monitoring task, the bit-rate classification task and the modulation format classification task are $y_i^{im} \in \mathbb{R}^3$ (three values of OSNR, CD and DGD, which are normalized to [0, 1]), $y_i^b \in \{0, 1\}$ (two bit-rates), and $y_i^m \in \{0, 1, 2\}$ (three modulation formats), respectively. Thus, the data label can be denoted as $\{\mathrm{y}_i^{im}, \mathrm{y}_i^b, \mathrm{y}_i^m\}_{i=1}^{N}$. As a regression problem, the loss function of the impairments monitoring task can be expressed as

$$loss_{im} = \frac{1}{2}\left\| y_i^{im} - f\left(x_i; W^{im}\right)\right\|^2 \tag{3}$$

where W^{im} is the network parameters of the impairments monitoring task. $f\left(x_i; W^{im}\right)$ is the output of the impairments monitoring task. As a classification problem, the loss function of the bit-rate classification is the cross-entropy, defined as

$$loss_{\mathrm{b}} = y_i^b log\left(p\left(y_i^b \left| x_i; W^b\right.\right)\right) \tag{4}$$

where W^b is the network parameters of the bit-rate classification task. Similarly, the loss function of the modulation format classification can be expressed as

$$loss_m = y_i^m log\left(p\left(y_i^m \left| x_i; W^m\right.\right)\right) \tag{5}$$

where W^m is the network parameters of the modulation format classification task. Finally, the total loss of the simultaneous OPM and BR-MFI using multi-task learning can be expressed as

$$loss_{total} = \underset{W^{im}, W^b, W^m}{argmin}\left(\sum_{i=1}^{N} loss_{im} + \sum_{i=1}^{N}\lambda_1 loss_b + \sum_{i=1}^{N}\lambda_2 loss_m\right) \tag{6}$$

where λ_1 and λ_2 are the importance factors of the bit-rate classification task and the modulation format classification task, respectively. The method of the gradient descent can be used to minimize the total loss function. The importance factors have significant impact on model performance as analyzed in [14]. To balance different tasks and attain good performance, we choose $\lambda_1 = 1.2$ and $\lambda_2 = 0.8$ in our work.

We train four different models to evaluate the performance. The first model named as "abc" has complete structure as shown in Fig. 3 ($j = 3$). The second model is named as "ac", whose "Conv_b" layer is removed ($j = 2$). Similarly, the third model without "Conv_a" layer is named as "bc" ($j = 2$). The last model without "Conv_a" and "Conv_b" layers is named as "c", which means no feature to fuse ($j = 1$). Noted that all the four models are created based on the idea of feature sharing but with different degrees of feature fusion. Specifically, the "abc" model fuses all features from the three levels, the "ac" model fuses the features from the low and high levels, the "bc" model fuses the features from the mid and high levels. The "c" model has no feature to fuse, which means the "c" model doesn't have the ability of feature fusion. The comparison among the four models can emerge the effect of feature fusion.

Fig. 5 shows several feature maps (reshaped from 100×1 to 10×10) of the "FC_2" layer in all models. As demonstrated by the confident scores (red bars under each image) and the similar feature maps generated by one model under the same attribute, the three models with the capability of

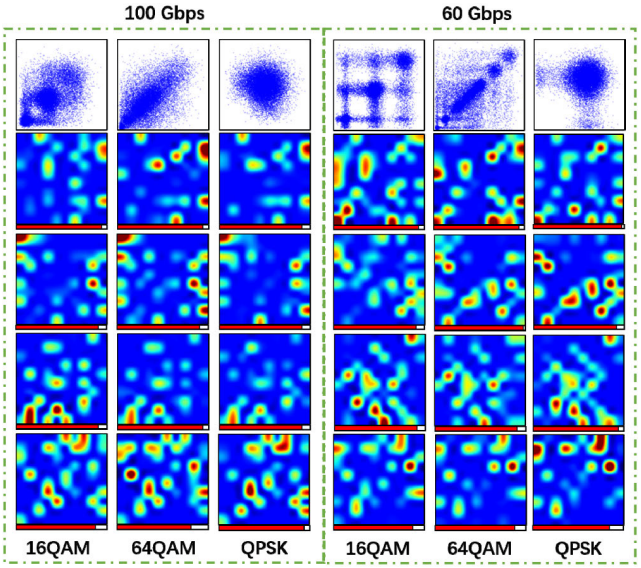


FIGURE 5. The feature maps reshaped from the "FC_2" layer of the "abc" model (2nd row), the "bc" model (3rd row), the "ac" model (4th row), and the "c" model (5th row) in response to the phase portraits (1st row). The red bar under each feature map is the confident score, where the higher score denotes higher probability to be the right modulation format.

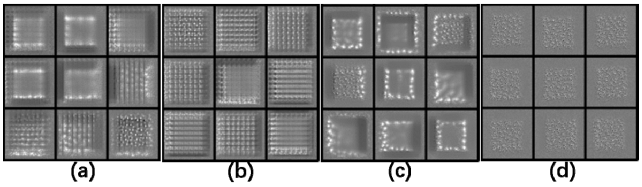


FIGURE 6. Learned filters from the "Conv_all" layer of (a) the "abc" model, (b) the "ac" model, (c) the "bc" model and (d) the "c" model.

feature fusion yield relatively higher confident scores than "c" model. Moreover, the "abc" model outperforms "ac" and "bc" models, "ac" and "bc" model yield close confident scores. We can conclude form the similar feature maps generated by single model that all the four models have the capability of feature sharing. Moreover, the differences among feature maps generated by each model show that, the different ways of feature fusion lead to different optimization spaces, which can be used to improve model performance.

Then, we present a part of filters in the "Conv_all" layer learned by the four models. As shown in Fig. 6, the filters learned by "c" model are visually flat and smooth, which means they mainly capture high-level features. When the mid-level feature is fused with the high-level feature, the filters learned by "bc" model appear to have slight geometric structure around and flat centrally. It means that the filters learned by "bc" model are trying to capture information from both the mid-level and high-level features. When the low-level feature is fused with the high-level feature, the filters learned by "ac" appear strong geometric structure and very slight flat part. It means that the filters learned by "ac" model are trying to capture information from both the low-level and high-level features. Finally, when the low-level, mid-level, and high-level features are fused together, the filters learned

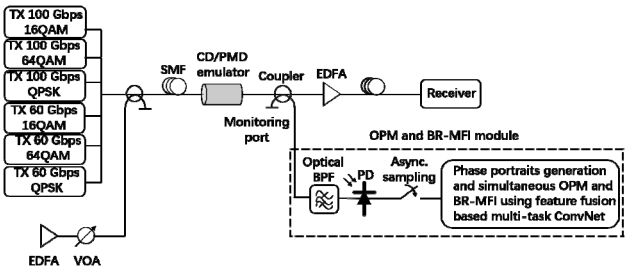


FIGURE 7. Simulation system for simultaneous OPM and BR-MFI by feature fusion based multi-task ConvNet. In the OPM and BR-MFI module, the sampled signal sequence is converted to the phase portraits by the technique of ASCS, then the feature fusion based multi-task ConvNet is used to predict results from the phase portraits.

by "abc" model present all the patterns in the filters of other three models, which means they are capturing information from the three levels of features. With the help of feature fusion, the model can form a new optimization space, which may improve the model performance.

III. SYSTEM SETUP AND RESULTS

The numerical simulation system is established using the Virtual Photonics Inc. (VPI) software and Tensorflow library [21], [22]. Fig. 7 shows the schematic diagram of the system configuration. In the transmitter, six different signals (60/100 Gbps QPSK, 60/100 Gbps 16QAM, 60/100 Gbps 64QAM) are generated and then transmitted over a single-mode fiber (SMF). To simulate the value of optical impairments, three important components are used. The first component which is used to adjust OSNR in the range of 10-28 dB (in steps of 2 dB) is consisted of the erbium-doped fiber amplifier (EDFA) and variable optical attenuator (VOA). The second component is the CD emulator. It is used to add CD in the range of 0-450 ps/nm (in steps of 50 ps/nm). The DGDs of the signals are varied in the range of 0-10 ps (in steps of 1 ps) by using the third component which is the polarization-mode dispersion (PMD) emulator. Noted that for each value of DGD, the angle α of the PMD emulator between the principal states of polarization (PSP) and the signal's state of polarization (SOP) is varied at random.

When the system starts to simulate, we use the optical coupler to extract a fraction of fixed power (-6 dBm) optical signal, then the extracted signal is sent into the OPM and BR-MFI module for subsequent the processing. First, the tapped signal is filtered to get the desired channel by an optical band-pass filter (BPF), then the filtered signal is directly detected by the PD. The electrical and optical bandwidths of the PD are 50 GHz and 0.8 nm, respectively. The electrical signal converted from optical signal is sampled asynchronously with a slow rate to collect a sequence of 30,000 samples. Next, the sample pairs (S_i, S_{i+k}) with fixed sample spacing k of 10 are acquired. Finally, the phase portraits (in ".png" format) are generated by the acquired sample pairs. We collect a dataset of 6600 phase portraits correlated with 2 bit-rates, 3 modulation formats, 10 CDs, 10 OSNRs, and 11DGDs. The bit-rate label vectors (60 Gbps:

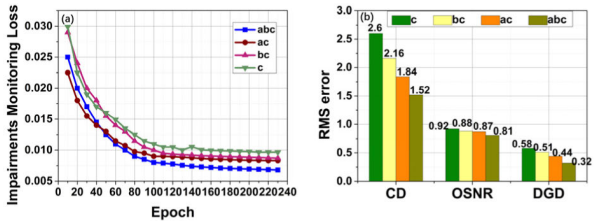


FIGURE 8. (a) The impairments monitoring losses in response to the epochs for four models. (b) The RMS errors of three impairments for four models.

10, 100 Gbps: 01) represent two different bit-rates. Similarly, the format label vectors (QPSK: 100, 16QAM: 010, 64QAM: 001) represent three different modulation formats. Two separate subsets named as training and testing set are randomly divided from the original dataset according to the proportion of 0.9 and 0.1, respectively. During the process of training, our feature fusion based multi-task ConvNet can improve model performance gradually by the capability of feature extraction, sharing, and fusion. The testing set is used to investigate the model performance.

A. IMPAIRMENTS MONITORING

Firstly, the impairments monitoring losses under different epochs for the four models are shown in Fig. 8(a). It is evident that the losses reduce with the increase of the epochs for all models. Moreover, the ''abc'' model with low-level, mid-level, and high-level feature fusion outperforms the others, and the performance of the ''ac'' model is better than the ''bc'' model's. The worst performance is achieved by the ''c'' model. It is because that the ''c'' model only use the high-level feature, which contains less impairments information than the fused features. The RMS errors over different impairments of the four models are shown in Fig. 8(b). The RMS errors are 1.52 ps/nm, 0.81 dB, and 0.32 ps for the ''abc'' model, which outperforms the others.

Moreover, the detailed performance of the ''abc'' model is shown in Fig. 9. The impairments monitoring results against the true values are shown in Fig. 9(a)-(c). It can be seen that the estimations of the CD, OSNR, and DGD are quite accurate. The ''abc'' model's RMS errors of six signals under each impairment values are shown in Fig. 9(d)-(e) in detail.

It is noted that in Fig. 9(d), the three signals of 100 Gbps have larger errors than the other three signals of 60 Gbps. And for the signals under the same rate, higher order modulation signals are more difficult to estimate. This is because that the signals with higher bite-rate and higher order format are more susceptible to the CD impairment, causing the greater estimation errors. It is clear that the trend of the CD estimation error decrease with the increase of the reference CD. It is because that when CD starts to increase in the monitoring range (0-450 ps/nm), each step of the growth can bring notable distinguish on phase portraits, which is beneficial for the monitoring.

It can be seen from Fig. 9(e) that in the range of 10-20 dB, the errors of all six signals decrease with the increase

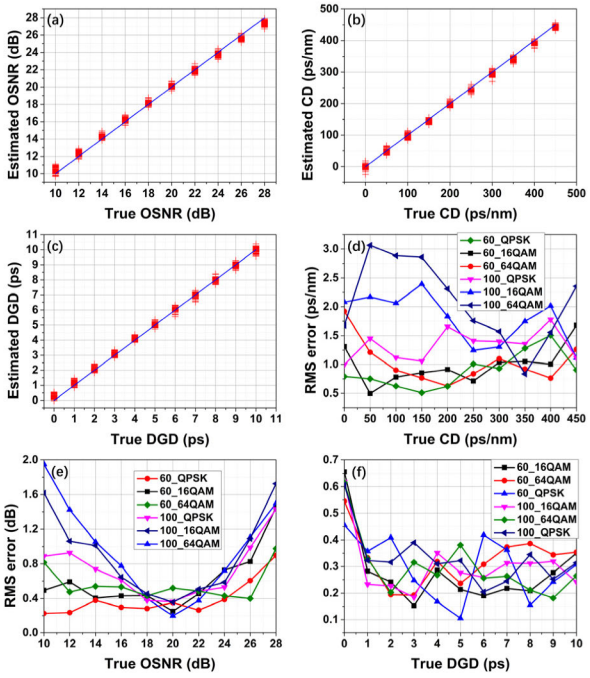


FIGURE 9. Impairments monitoring results of the ''abc'' model against true values of (a) OSNR, (b) CD, (c) DGD. The RMS errors of the ''abc'' model for each (d) CD, (e) OSNR, (f) DGD for all six signals.

of OSNR. However, in the range of 20-28 dB, they increase with the increase of OSNR. The explanation of the reduction of the OSNR estimate errors is similar to the reduction of the CD estimate errors'. However, different with the monitoring range of CD, the monitoring range of OSNR contains a threshold (20 dB). When OSNR surpass the threshold, the OSNR effect on phase portraits is becoming weaker and weaker, which is difficult for the model to estimate.

Fig. 9(f) shows the RMS errors of DGD for all signals. It is obvious that when the DGD equals to zero, the errors of all six signals are relatively large. However, when DGD reaches the value of 1 and increases in the range of 1-10 ps, the RMS errors drop suddenly and oscillates around 0.28 ps. The reason is that when DGD equals to zero, no DGD effect is added to phase portraits, which is difficult for the model to estimate. When DGD changes in the range of 1-10 ps, the adjacent DGD values have the similar effect on phase portraits. The results show that our feature fusion based multi-task model can monitor impairments of large range precisely.

B. BIT-RATE AND MODULATION FORMAT IDENTIFICATION

Secondly, we investigate the BR-MFI performance of the four models as shown in Fig. 10. The name of each curve consists of the abbreviation for the specific classification task and the name of the model. For instance, the name ''br_c'' means that the curve represents the bit-rate classification accuracies of the ''c'' model. Similarly, the name ''mfi_c'' means that the curve represents the modulation format classification accuracies of the ''c'' model. Noted that the accuracies of bit-rate classification and modulation format classification

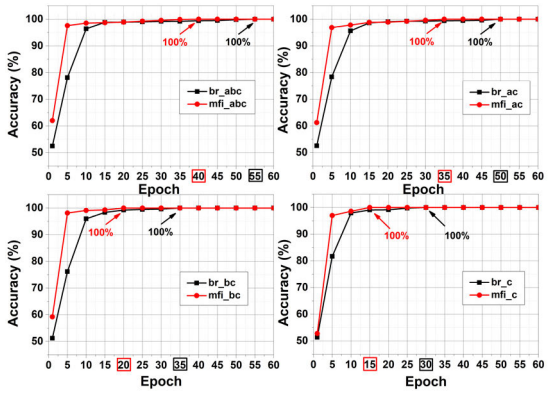


FIGURE 10. The classification accuracy curves for BR-MFI of (a) the "abc" model, (b) the "ac" model, (c) the "bc" model and (d) the "c" model. The red curve represent the modulation format identification, the black curve represent the bit-rate identification.

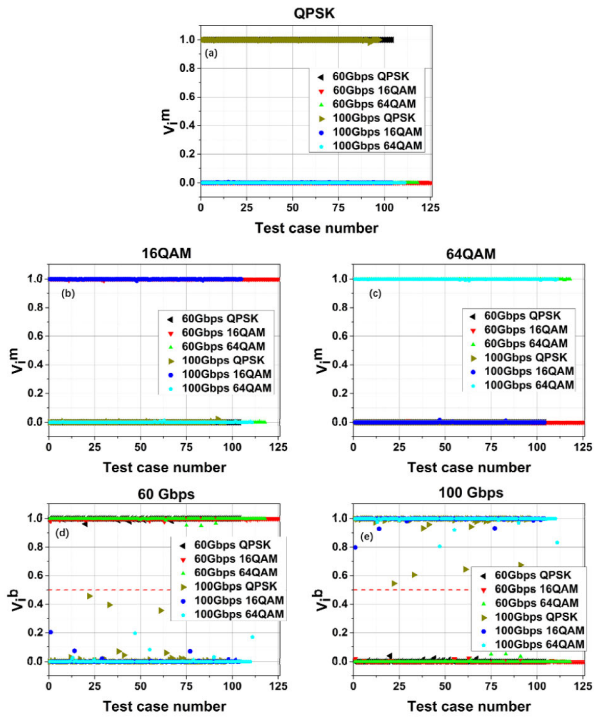


FIGURE 11. Elements of the "abc" model output vectors v_i^m and v_i^b for (a) QPSK, (b)16QAM, (c) 64QAM and (d) 60 Gbps, (e) 100 Gbps, respectively in response to the phase portraits in the testing set.

all attain 100% for each model. Specifically, the minimum epochs to attain 100% accuracy of bit-rate classification and format classification for "abc", "ac", "bc", and "c" models are (55,40), (50,35), (35,20), and (30,15), respectively. We can conclude that for each model, the bit-rate classification task spends larger epoch than the format classification task. This is because that format can present more distinct phase portraits than bit-rate, which is easier to recognize. Moreover, it is found that the "abc" model spend larger training epochs than the "ac", "bc" and "c" models, and the "c" model spend the minimum training epochs. The reason for this is that the low-level and mid-level features in the "abc" model contain many visual information, which is not

TABLE 1. The used filter size and stride of each layer to produce the same shape feature maps.

Layer	1	2	3
Filter size & Stride	Size:(21,21) Stride:10	Size:(19,19) Stride:10	Size:(11,11) Stride:5
Layer	4	5	6
Filter size & Stride	Size:(11,11) Stride:5	Size:(9,9) Stride:2	Size:(7,7) Stride:2
Layer	7	8	
Filter size & Stride	Size:(5,5) Stride:2	Size:(3,3) Stride:2	

TABLE 2. Model performance in response to the combination of locations.

	Layer 1	Layer 2	Layer 3
Layer 4	(1) 0.0153 (2) 8 (3) 0.0521	(1) 0.0136 (2) 5 (3) 0.0565	(1) 0.0124 (2) 6 (3) 0.0537
Layer 5	(1) 0.0149 (2) 8 (3) 0.0510	(1) 0.0142 (2) 6 (3) 0.0555	(1) 0.0137 (2) 4 (3) 0.0519
Layer 6	(1) 0.0092 (2) 3 (3) 0.0507	(1) 0.0074 (2) 2 (3) 0.00546	**(1) 0.0068** **(2) 0** **(3) 0.0513**
Layer 7	(1) 0.0116 (2) 3 (3) 0.0498	(1) 0.0113 (2) 2 (3) 0.0531	(1) 0.0112 (2) 2 (3) 0.0512

(1) Impairments monitoring loss
(2) Number of bit-rate classification error
(3) Test time per image (s)

suitable for classification task. These visual information need to be encoded and extracted to form the concept of class, thus it would spend more training epochs.

Then, Fig. 11 shows the elements of the two output classification vectors v_i^b and v_i^m of the "abc" model in response to the input phase portraits from the testing set. The vector v_i^b has two neurons for the bit-rate classification task. The vector v_i^m has three neurons for the modulation format classification task. Each element of v_i^b and v_i^m represents the probability of the distinct bit-rate and format type, respectively. Moreover, the sum of the probabilities of the two output neurons of the bit-rate classification equals to 1. Similarly, the sum of the probabilities of the three output neurons of the format classification task also equals to 1. The $argmax\left(v_i^m\right)$ as well as $argmax\left(v_i^b\right)$ is taken as the classified modulation format and bit-rate, respectively. For modulation format classification task, the huge separation between the largest element in v_i^m and the rest means that the performance of the modulation format classification is excellent. However, the separation between the elements of v_i^b are relatively small, and several 100 Gbps QPSK signals are very easy to be misidentified to the opposite rate, since they are close to the classification threshold of 0.5. Again, the above results confirm that the bit-rate classification task is more difficult than the modulation format classification task.

C. THE LOCATION AND SHAPE OF EXTRACTED FEATURE FOR FUSION

Besides, we study the influence of different location and shapes of the extracted features on the "abc"

model performance. Firstly, the influence of the location is studied. As mentioned above, the features are distributed hierarchically, and the adjoining layers are highly relevant, it is no need to consider the fusion of all the adjoining layers. Moreover, Table 1 shows the filter size and filter stride of each layer in ''Backbone'' part that we used to obtain the feature map with the fixed dimension of $8 \times 8 \times 256$ for fusion. We extract features from the low-level part (1-3 layers) and mid-level part (4-7 layers) with different combination of locations, then concatenate them with the high-level feature (8th layer) to train twelve models. Three indicators (impairment monitoring loss, number of bit-rate classification error, and test time per image) reflecting model performance in response to the different combination of location are shown in Table 2. Noted that all the twelve models attain the accuracy of 100% on the modulation format classification task, so it is ignored in Table 2.

It is found that when the location of feature extraction varies among the mid-level part (4-7 layers), the impairments monitoring losses of the fourth and fifth layers are close. The sixth layer obtains the minimum loss. When the location of extracted feature varies among the low-level part (1-3 layer), the impairments monitoring loss decreases with the increase of the layer number. This is because that the network structure behind the fusion layer is fixed and relatively simple, thus its capability of feature extraction is limited. Instead of using the features which are relatively raw from the front part of the ''Backbone'', the features from the latter part which are fully processed is easier to get lower impairment monitoring loss.

For the number of the bit-rate classification error, when the location of feature extraction varies among the mid-level part (4-7 layer), the sixth layer obtains the minimum number of errors. It is because that the seventh and eighth layer are adjacent layers, the features of them are highly correlated and similar. Therefore, the fusion of features from the seventh and eighth layers is helpless to improve the performance of classification. In addition, the differences of test time among all models are small, and all models satisfy the requirement of real-time monitoring ($\sim$ 51 ms). Overall, when the features are extracted from the third and sixth layer, the fused feature can help the model to get optimal results (minimum loss, no error, test time is $\sim$51 ms on Intel Core i7 CPU).

Then, the influence of the feature shape is investigated. In this part, we mainly consider the impairments monitoring task. The shape of the feature consists of the width, the height, and the number of channels. In particular, in this work, the width is equal to the height, thus the length of the edge is used to describe them. We adjust the number of channels and the length of the edge when other parameters are fixed. The number of channels is changed in the range of 128-384 (in steps of 64), and the length of the edge is changed in the range of 4-12 (in steps of 2). The impairments monitoring losses in response to the different feature shapes are shown in Fig. 12. It is clear that when the number of channels and the length of edge are inappropriate (too large or too small), the losses would increase. The minimum loss is achieved

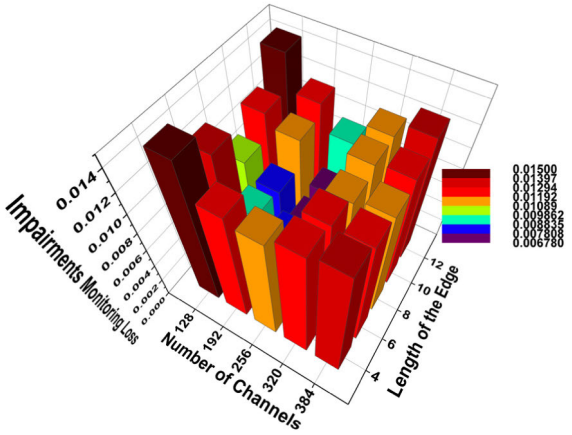


FIGURE 12. The impairments monitoring losses in response to the feature shape.

when the number of channels and the length of edge are 256 and 8, respectively. It is because that the quantity of useful information is limited, more channels or larger edge lead to huge amount of parameters, and less channels or smaller edge cannot obtain all useful information, resulting in the suboptimal results.

As analyzed above, we can conclude that the location and shape of extracted feature for fusion do have significant influence on models' performance. It is crucial to select appropriate location and feature shape for fusion. For the task of simultaneous OPM and BR-MFI, we think that the location of (layer 3, layer 6) and the feature shape of $8 \times 8 \times 256$ is a suitable setting.

D. FIBER NONLINEARITY AND TRANSMITTER VARIATIONS

Next, in order to analyze the reliability of the proposed technique, fiber nonlinearity and transmitter variations are considered. Specifically, due to the manufacturing tolerances and components ageing of the transmitter, the rise and fall times of the pulses as well as the extinction ratios would change slightly. Moreover, in practical optical networks, the influence of the fiber nonlinearity cannot be ignored. Therefore, the two factors would affect the performance of our technique. In particular, one transmitter which is dissimilar to the ideal transmitter (without transmitter variations) is used to generate new testing set. The rise/fall times of the pulses and extinction ratios generated by this transmitter and the ideal transmitter differ in the range of 15%-25% and 3 dB, respectively. On the other hand, we change the system configuration of fiber nonlinear coefficient γ and input power from 0/W$\cdot$ km and -6 dBm to 1.2/W$\cdot$ km and 0 dBm, respectively. The transmission fiber link is 1000 km. The changed system is used to generate another new testing set which is affected by the fiber nonlinear effects. Finally, the ''abc'' model trained by the original training set (generated by ideal transmitter and without the effect of fiber nonlinearity) is used to evaluate above two new testing sets. Fig. 13 and Fig. 14 show the results of impairments monitoring and BR-MFI for two scenarios (transmitter variations and fiber nonlinearity).

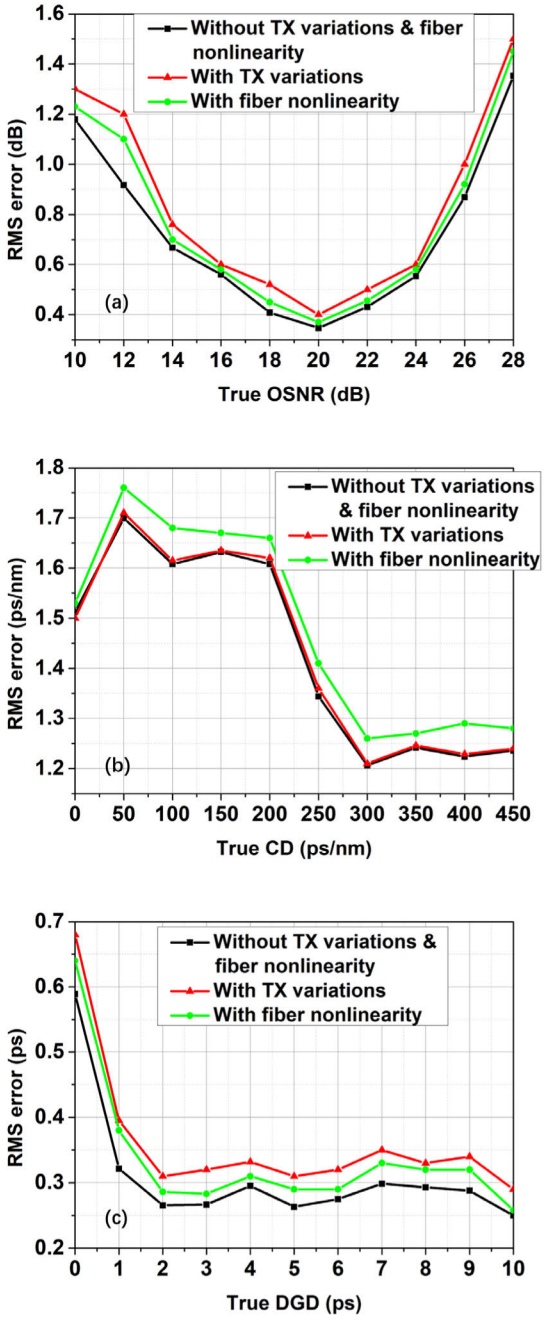


FIGURE 13. The RMS errors of (a) OSNR, (b) CD and (c) DGD in response to transmitter variations and fiber nonlinearity, respectively.

As for the transmitter variations, it is obvious that the RMS errors of the CD estimation remain almost unaffected, and the RMS errors of the OSNR and DGD are increased by ∼0.1 dB and ∼0.06 ps, respectively. Affected by transmitter variations, the accuracy of the BR-MFI drops from 100% to 99.7%, as shown in Fig. 14(a). Specifically, two 100 Gbps QPSK phase portraits are misclassified as 60 Gbps QPSK phase portraits. On the other hand, for the fiber nonlinearity, it is clear that the RMS errors of the CD, OSNR and DGD are increased by ∼0.057 ps/nm, ∼0.030 dB and ∼0.032 ps. As shown in Fig. 14(b), the overall accuracy drops from

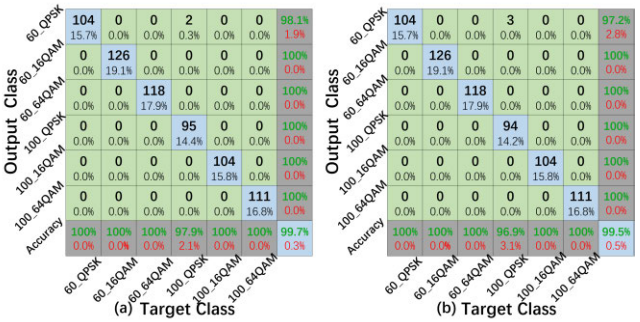


FIGURE 14. Confusion matrix for BR-MFI under the scenarios of (a) transmitter variations and (b) fiber nonlinearity.

100% to 99.5% since three 100 Gbps QPSK phase portraits are misclassified as 60 Gbps QPSK phase portraits. We can conclude that our feature fusion based multi-task ConvNet is robust against the transmitter variations and fiber nonlinearity. Moreover, in order to improve the robustness of our technique, the phase portraits which are affected by the transmitter variations and the fiber nonlinearity should be used in the training of the model with the convenient method of transfer learning [23].

IV. CONCLUSION

In this paper, we propose a feature fusion based multi-task ConvNet for simultaneous OPM and BR-MFI. Instead of training the multi-task ConvNet only using the feature from the last layer, the feature differences of different layers are considered. The fused feature of low-level, mid-level, and high-level is used to train the multi-task ConvNet to improve the model performance. With the proposed method, the RMS errors of OSNR, CD, and DGD are 0.81 dB, 1.52 ps/nm, and 0.32 ps, respectively. Besides, 100% accuracy is attained for the BR-MFI task. Moreover, the effects of the location, the shape of feature extracted for fusion, the transmitter variations, and the fiber nonlinearity on model performance are also investigated. Due to the rapid running time (∼51 ms) and high monitoring precision, this technique is a promising candidate for simultaneous OPM and BR-MFI in future dynamic optical networks.

REFERENCES

[1] D. C. Kilper, R. Bach, D. J. Blumenthal, D. Einstein, T. Landolsi, L. Ostar, M. Preiss, and A. E. Willner, "Optical performance monitoring," *J. Lightw. Technol.*, vol. 22, no. 1, pp. 294–304, Jan. 15, 2004.

[2] Z. Pana, C. Yub, and A. E. Willner, "Optical performance monitoring for the next generation optical communication networks," *Opt. Fiber Technol.*, vol. 16, no. 1, pp. 20–45, Jan. 2010.

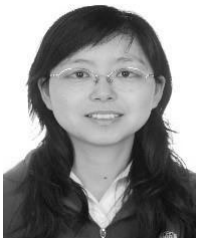
[3] F. N. Khan, T. S. R. Shen, Y. Zhou, A. P. T. Lau, and C. Lu, "Optical performance monitoring using artificial neural networks trained with empirical moments of asynchronously sampled signal amplitudes," *IEEE Photon. Technol. Lett.*, vol. 24, no. 12, pp. 982–984, Jun. 2012.

[4] D. Wang, M. Zhang, Z. Cai, Y. Cui, Z. Li, H. Han, M. Fu, and B. Luo, "Combatting nonlinear phase noise in coherent optical systems with an optimized decision processor based on machine learning," *Opt. Commun.*, vol. 369, pp. 199–208, Jun. 2016.

[5] D. Wang, M. Zhang, M. Fu, Z. Cai, Z. Li, H. Han, Y. Cui, and B. Luo, "Nonlinearity mitigation using a machine learning detector based on *k*-nearest neighbors," *IEEE Photon. Technol. Lett.*, vol. 28, no. 19, pp. 2102–2105, Oct. 1, 2016.

[6] F. N. Khan, Y. Zhou, Q. Sui, and A. P. T. Lau, "Non-data-aided joint bit-rate and modulation format identification for next-generation heterogeneous optical networks," *Opt. Fiber Technol.*, vol. 20, no. 2, pp. 68–74, 2014.
[7] 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*, vol. 23, no. 23, pp. 30337–30346, Nov. 2015.
[8] 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*, vol. 25, no. 15, pp. 17767–17776, Jul. 2017.
[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[10] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, Jun. 2014.
[11] T. N. Sainath, B. Kingsbury, G. Saon, H. Soltau, A.-R. Mohamed, G. Dahl, and B. Ramabhadran, "Deep convolutional neural networks for large-scale speech tasks," *Neural Netw.*, vol. 64, pp. 39–48, Apr. 2015.
[12] D. Wang, M. Zhang, Z. Li, J. Li, M. Fu, Y. Cui, and X. Chen, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photon. Technol. Lett.*, vol. 29, no. 19, pp. 1667–1670, Oct. 2017.
[13] D. Wang, M. Zhang, J. Li, Z. Li, J. Li, C. Song, and X. Chen, "Intelligent constellation diagram analyzer using convolutional neural network-based deep learning," *Opt. Express*, vol. 25, no. 15, pp. 17150–17166, 2017.
[14] 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 Photon. J.*, vol. 10, no. 5, Oct. 2018, Art. no. 7906712.
[15] Y. Zhang and Q. Yang, "An overview of multi-task learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 30–43, Jan. 2018.
[16] Z. Wu, C. Valentini-Botinhao, O. Watts, and S. King, "Deep neural networks employing multi-task learning and stacked bottleneck features for speech synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 4460–4464.
[17] S. D. Dods and T. B. Anderson, "Optical performance monitoring technique using delay tap asynchronous waveform sampling," in *Proc. Opt. Fiber Commun. Conf. Nat. Fiber Optic Eng. Conf.*, Mar. 2006, pp. 175–192.
[18] Y. Yu and C. Yu, "OSNR monitoring by using single sampling channel generated 2-D phase portrait," in *Proc. Opt. Fiber Commun. Conf. Exhibit.*, Mar. 2014, pp. 1–3.
[19] Y. Yu and C. Yu, "Optical signal to noise ratio monitoring using variable phase difference phase portrait with software synchronization," *Opt. Express*, vol. 23, no. 9, pp. 11284–11289, 2015.
[20] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2013, pp. 818–833.
[21] VPIsystems, "VPItransmissionMaker," VPIphotonics, Berlin, Germany, 2009.
[22] M. Abadi, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," in *Proc. Conf. Lang. Resour. Eval.*, Mar. 2016, pp. 3243–3249.
[23] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

XIAOJIE FAN received the bachelor's degree in communication engineering from the University of Science and Technology Beijing, Beijing, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Computer and Communication Engineering. His research interests include optical communication and deep learning.



LINA WANG received the M.S. and Ph.D. degrees in communication and information systems from the Harbin Institute of Technology, in 2001 and 2004, respectively. From 2010 to 2012, she was a Postdoctoral Research with the University of Science and Technology Beijing, where she is currently an Associate Professor with the Department of Communication Engineering, School of Computer and Communication Engineering. Her research interests include space communications, BeiDou satellite positioning algorithms, fountain codes, optical communication systems, optical devices, deep learning, network security, and game theory.

FANG REN received the bachelor's and master's degrees in electronic science and technology from Tianjin University, Tianjin, China, in 2008 and 2010, respectively, and the Ph.D. degree in information electronics from Hokkaido University, Sapporo, Japan, in 2014. Her research interests include optical communication systems, optical devices, and deep learning.

YULAI XIE received the bachelor's and master's degrees in electronic science and technology from Tianjin University, Tianjin, China, in 2008 and 2010, respectively, and the Ph.D. degree in information electronics from Hokkaido University, Sapporo, Japan, in 2014. He is currently a Researcher with Hitachi (China) Research and Development Company Ltd., Beijing, China. His research interest includes deep learning.

XIANG LU received the bachelor's degree in computer science and technology from Shanxi Normal University, Xi'an, China, in 2007, and the master's degree in communication and information system from Guizhou University, Guiyang, China, in 2012. He is currently a Researcher with the Information Center, Guizhou Power Grid Company Ltd., Guiyang. His research interest includes devops.

YIYING ZHANG received the bachelor's degree in communication engineering from the University of Science and Technology Beijing, Beijing, China, in 2015, where she is currently pursuing the Ph.D. degree with the School of Computer and Communication Engineering. Her research interest includes optical communications.

TIANWEN ZHANGSUN received the bachelor's degree in communication engineering from the University of Science and Technology Beijing, Beijing, China, in 2017, where she is currently pursuing the master's degree with the School of Computer and Communication Engineering. Her research interest includes optical communications.

WEI CHEN received the bachelor's degree in communication engineering from the University of Science and Technology Beijing, Beijing, China, in 2017, where he is currently pursuing the master's degree with the School of Computer and Communication Engineering. His research interests include deep learning and robotics.

JIANPING WANG received the bachelor's, master's, and Ph.D. degrees from the School of Precision Instrument and Optoelectronic Engineering, Tianjin University, Tianjin, China, in 1995, 1997, and 2000, respectively. Her research interests include optical communications, microwave photonics, and deep learning.