Feedforward and Recurrent Neural Network-Based Transfer Learning for Nonlinear Equalization in Short-Reach Optical Links

Zhaopeng Xu ⬤, Chuanbowen Sun ⬤, Tonghui Ji, Jonathan H. Manton ⬤, *Fellow, IEEE*, and William Shieh ⬤, *Fellow, IEEE*

Abstract—**Neural network (NN)-based nonlinear equalizers have been shown effective for various types of short-reach direct detection systems. However, they work best for a certain channel condition and need to be trained again when the channel environment is changed, which hinders the efficient deployment of future optical switched data center networks. In this article, we propose transfer learning (TL)-aided feedforward neural networks (FNN) and recurrent neural networks (RNN) for nonlinear equalization in short-reach direct detection optical links, which enables a fast transition to new equalizers when the channel condition is changed. A 50-Gb/s 20-km pulse amplitude modulation (PAM)-4 optical link is experimentally demonstrated as the target system, and links of varying bit-rates and fiber lengths are selected as the source system. Experimental results show that TL could help reduce the number of epochs and training symbols of FNNs/RNNs required for nonlinear equalization in the target system, taking advantage of FNNs/RNNs trained for source systems. A reduction of 90%/87.5% in epochs and 62.5%/53.8% in training symbols is achieved with FNNs/RNNs transferred from the most similar source system. We also find that FNNs can be transferred to their corresponding RNNs for equalization in the target system, while TL from RNNs to FNNs cannot work properly. TL enables a fast transition between different NN-based equalizers, which is critical for future optical switched data center networks, where the optical links need to be dynamically reconfigured.**

Index Terms—**Direct detection, neural networks, nonlinearity, transfer learning.**

I. INTRODUCTION

WITH the exponential growth of the Internet traffic, there has been an increasing demand for bandwidth-hungry applications such as data center interconnects (DCI), which results in the rapid development of high-speed short-reach optical

links [1]–[4]. Instead of coherent detection, direct detection is widely adopted for short-reach applications to meet the low-cost requirement, and cost-effective devices such as the directly modulated lasers (DML) are commonly used to further reduce expenditure [5], [6]. However, the reduction of system structure and cost would inevitably bring in many impairments, for example, the frequency chirp of the DMLs, the non-ideal frequency response of devices, and most importantly, the nonlinearity induced by intensity direct detection mixed with linear effects such as chromatic dispersion (CD) [7], [8]. These effects will introduce severe inter-symbol interference (ISI) to short-reach direct detection systems, and it is hard to fully mitigate channel effects with intensity-only detection. Efficient nonlinear equalization techniques are of vital importance to guarantee a desired system performance.

With the rapid development of computing devices, machine learning (ML) has become an extremely hot topic in recent years. Many ML algorithms have been applied for optical fiber communications, and for nonlinear equalization in short-reach optical links, different types of neural network (NN)-based nonlinear equalizers [9]–[19] have been widely investigated and regarded as one of the most promising solutions. These NN-based nonlinear equalizers could be potentially implemented for real time applications [20], with superior performance compared with traditional high-order Volterra series-based nonlinear equalizers. With the help of nonlinear activation function in each hidden layer, NNs are well suited to solve this kind of nonlinear approximation problems in short-reach direct detection links. However, before serving as efficient equalizers, NNs need to be trained with a larger number of training symbols after a large number of epochs, which is highly computationally inefficient. What's more, a certain NN-based equalizer can only work best for one optical link. In other words, when the link scenario is changed, e.g., different bit-rates or fiber lengths, the performance of the old NNs will degrade and the NNs may need to be retrained to fit for the new scenario. In [21], an end-to-end deep learning scheme is proposed, and the training process can be done over multiple links so that the equalizer can be used for a range of scenarios. This approach eliminates the need to retrain the network, however, it may sacrifice some performance since different set of data are used for training compared with training on one specific data set. Moreover, the trained NN may not able to efficiently handle scenarios outside the training data

sets. Therefore, study of retrained NNs for each link is still of great interest.

In this paper, we extend our previous work [22] and propose transfer learning (TL)-aided feedforward NNs (FNNs) and recurrent NNs (RNNs) for nonlinear equalization in short-reach direct detection optical links. Compared with [22], RNN-based TL and the conversion between the two types of NNs are taken into account. Besides, in addition to the individual effects of source systems such as bit-rate or fiber length discussed in [22], this paper considers changes of both effects. TL is commonly used for image processing, where the information of one set of images could be "transferred" to a different set using part of trained NNs [23], and it has been introduced to optical communication for optical performance monitoring [24]–[26] or nonlinear mitigation in medium to long-haul optical single side-band (SSB) systems [27]. For the nonlinear equalization purpose of double side-band (DSB) short-reach direct detection systems in this paper, TL can make use of the knowledge gained from NNs trained for one short-reach link (source system), and apply it to a different but related link (target system). This process reduces the number of training symbols and epochs significantly, since the NNs for the target system are now trained with preserved information gained from the source systems. We experimentally demonstrated a 50-Gb/s 20-km pulse amplitude modulation (PAM)-4 direct detection link as the target system, and links of different bit-rates and fiber lengths are selected as the source ones. The system nonlinearity is mainly induced by the mixture of linear impairments (such as CD and bandlimited effects) with square-law detection, the DML frequency chirp, and the device nonlinearity. Both FNNs and RNNs are tested to be applicable for TL, where about 90%/87.5% reduction in epochs and 62.5%/53.8% in training symbols are achieved with FNNs/RNNs transferred from the most similar source system (60-Gb/s 15-km). TL of FNNs and RNNs are also performed with varying received optical power (ROP), and we show that NNs trained at high ROP work well even without retraining, since the FNNs or RNNs are obtained with less noise. Furthermore, we demonstrate that FNNs can be transferred to their corresponding RNNs for equalization in the target system, while TL from RNNs to FNNs fail to work properly. TL-aided NNs can facilitate the establishment of an optical link connection in an expeditious manner, and the fast equalizers may be critical for future optical switched data center networks, where the optical interconnects need to be dynamically reconfigured.

II. PRINCIPLE OF FNN- AND RNN-BASED NONLINEAR EQUALIZERS AND TL

Both FNN- and RNN-based nonlinear equalizers are applicable for TL, and their structures are shown in Fig. 1. Note that the RNN employed here refers to the autoregressive RNN (AR-RNN), since it demonstrates better system performance and lower computational complexity than the layer-RNN (L-RNN) on nonlinear equalization for short-reach applications. FNN and RNN are adopted in this paper since they have demonstrated both good system performance and moderate computational complexity [20] compared with other sophisticated NN variations.
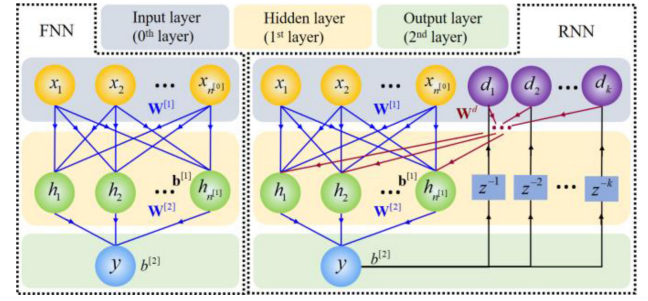

Fig. 1. Schematic of a 2-layer FNN and a 2-layer RNN.

For the sake of simplicity, we only consider 2-layer NN-based equalizers. For FNN, the i-th layer contains $n^{[i]}$ neurons, where $i = 0, 1, 2$. Only one output is used to predict the current PAM4 symbol level which reduces computational complexity. For the i-th layer, the weight matrix $\mathbf{W}^{[i]}$ is an $n^{[i]} \times n^{[i-1]}$ matrix containing all the weights connected from the (i-1)-th layer to the i-th layer, and the biases is represented by an $n^{[i]} \times 1$ vector $\mathbf{b}^{[i]}$. Assuming the activation function of the i-th layer is denoted by $f^{[i]}(\cdot)$, the one-time forward propagation (FP) of FNN, also referred to the symbol recovery process, is given by

$$y = f^{[2]}(\mathbf{W}^{[2]} f^{[1]}(\mathbf{W}^{[1]}\mathbf{x} + \mathbf{b}^{[1]}) + b^{[2]}), \quad (1)$$

where $\mathbf{x}$ is the input vector containing $n^{[0]}$ inputs and y is the output. For RNN, many output delays are involved based on the existing FNN structure, and these delays are connected to the hidden layer to pass on additional information. Assuming k delays are employed as feedbacks, the one-time FP of RNN can be expressed as

$$y = f^{[2]}(\mathbf{W}^{[2]} f^{[1]}([\mathbf{W}^{[1]}, \mathbf{W}^d][\mathbf{x}^T, \mathbf{y}_d^T]^T + \mathbf{b}^{[1]}) + b^{[2]}), \quad (2)$$

where $\mathbf{W}^d$ is an $n^{[i]} \times k$ matrix containing all the weights connected from the delays to the hidden layer (k denotes the number of the delays), $\mathbf{y}_d$ is an input containing a delayed version of the output sequence, and the superscript T denotes matrix transpose. RNN-based equalizers usually outperform FNN-based ones since the output delays in RNN introduce additional information, which are predictions of previous symbols at low bit-error-rate (BER).

The weights and biases of FNN or RNN are obtained in the training process, where a number of training symbols and epochs are needed. One epoch is defined as one-time back propagation (BP) over all the training symbols. Taking FNN as an example and assuming N_{train} training symbols are used, without the help of TL, the weights and biases are learned by solving the optimization problem shown as

$$\min_{\mathbf{W}^{[1]}, \mathbf{b}^{[1]}, \mathbf{W}^{[2]}, b^{[2]}} C(y_j, y_j^t), j = 1, 2, \ldots, N_{train}, \quad (3)$$

where $C(\cdot)$ denotes the cost function, the subscript j denotes the j-th training symbol, and the superscript t denotes the corresponding target symbol. The optimization problem is usually non-convex, and one simple solution to train the NNs is to perform stochastic gradient descent (SGD) through BP to approach a local minimum. The weights and biases are initialized
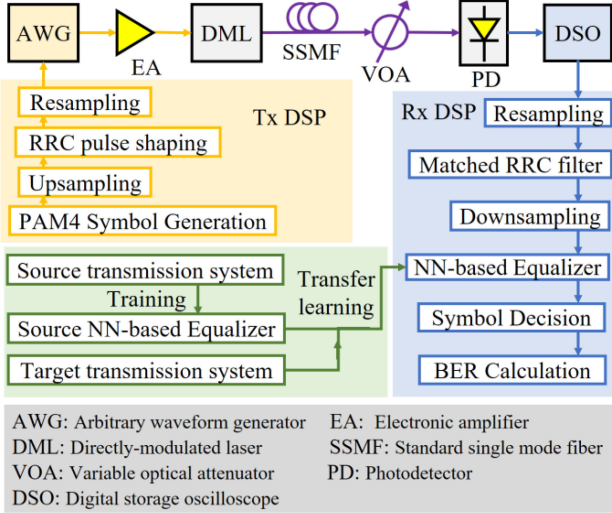
Fig. 2. Experimental setup of the PAM4 short-reach direct detection optical link using NN-based equalizers aided by TL.

randomly and get updated in each SGD step. Usually a large number of training symbols and epochs are needed for the cost function $C(\cdot)$ to come to a minimum point, which is rather computationally heavy and inefficient. TL could help speed up the training process (in the target system), making use of NNs trained for the source system. The TL-aided calculation of weights and biases of the target system can be shown as a new optimization problem described as

$$\min_{\mathbf{W}_T^{[1]}, \mathbf{b}_T^{[1]}, \mathbf{W}_T^{[2]}, b_T^{[2]}} C(y_j, y_j^t; \mathbf{W}_S^{[1]}, \mathbf{b}_S^{[1]}, \mathbf{W}_S^{[2]}, b_S^{[2]}),$$
$$j = 1, 2, \ldots, N_{train}, \tag{4}$$

where the subscript T denotes the target system, and S denotes the source system. By introducing prior information into the target system, we can simply set the initial weights and biases as the source ones, denoted by $\mathbf{W}_S^{[1]}, \mathbf{b}_S^{[1]}, \mathbf{W}_S^{[2]}, b_S^{[2]}$, to solve this optimization problem rather than choosing them all at random. Instead of training from scratch, the source weights and biases can serve as a better starting point for optimization in the target system. TL for RNN is similar to that of FNN by further including the influence of $\mathbf{W}^d$. As the NN size tends to be larger to better meet the high-speed low-cost requirement of short-reach transmission links, TL will play a more important role in future applications since the actual training time of NNs will become more significant compared with other factors such as the time for data collection and model deployment.

III. EXPERIMENTAL SETUP

The experimental setup of the DML-based PAM4 short-reach direct detection optical link is shown in Fig. 2, where TL-aided NNs are used for post nonlinear equalization. For the target system, the signal bit rate is 50 Gb/s, and the fiber length is 20 km. Links of different bit-rates (20/40/60/80 Gb/s) and fiber lengths (10/15/25/30 km) are selected as source systems. At the transmitter, the PAM4 signal are up-sampled and generated

by an arbitrary waveform generator (AWG, Keysight M8196A) operated at a sampling rate of 92 GSa/s. A root raised cosine (RRC) filter with a roll-off factor of 0.1 is used for Nyquist pulse shaping. An electrical amplifier with 17 dB gain (SHF 100 BP) is employed to amplify the electrical PAM4 signal and then the amplified signal is fed into a 16-GHz DML biased at 55 mA. The output power of DML is 9.5 dBm and standard single-mode fibers (SSMFs) of different lengths are used for transmission. At the receiver, a variable optical attenuator (VOA) is first employed to adjust the ROP and a 43-GHz photodetector (PD, BPDV2150R) is followed to directly detect the optical signal. The detected signal is then captured by a digital storage oscilloscope (DSO, DSA-X 93304Q) operated at a sample rate of 80 GSa/s. After resampling, matched RRC filtering and downsampling, the received PAM4 symbols are fed into TL-aided NN-based equalizers for nonlinear equalization. Finally, hard-decision is performed to the equalized signal, and the BER is calculated and collected.

As shown in the TL module in Fig. 2, NN-based equalizers for the target transmission system are obtained by applying TL to the NNs trained for the source transmission system. Both TL-aided FNNs and TL-aided RNNs are considered as the nonlinear equalizers, with 21 inputs consisting of the current symbol, 10 past symbols, and 10 post symbols. 12 hidden neurons are selected and 1 output is used as illustrated in Fig. 1. Compared with the 4-output NN structure, only 1 output is employed to reduce computational complexity. The 1-output NNs also provide more freedom to make concise additional connections onto FNNs. After comparison of different activation functions, the tanh function is chosen for the hidden layer due to its good performance and the pure-linear activation function is used for the output layer. For RNN-based equalizers, k is optimized and set as 2. All the hyperparameters of FNNs and RNNs are carefully chosen for good performance and fair comparison. Since the NNs have only one output to predict the PAM4 symbol level, mean square error (MSE) is selected as the cost function. After experimental comparison of different optimizers, Levenberg-Marquart (LM) algorithm is selected to carry out efficient update of network parameters for the training process. In addition to training symbols, 5000 different symbols are used in validation set to prevent overfitting while another 1.2 million PAM4 symbols are used for testing (MSE and BER calculation). Validation of the training is performed during the optimization process in each epoch.

IV. RESULTS AND DISCUSSIONS

The relationship of MSE and epochs for the target system is shown in Fig. 3 with retrained FNN-based equalizers from source systems. Three different types of source systems are chosen to prove the effectiveness of TL, where Fig. 3(a), 3(b), and 3(c) show source systems by varying data rates only, or varying fiber lengths only, or varying both bit-rates and fiber lengths, respectively. The number of training symbols is 4000, and the ROP is set at -1 dBm. Even though the bit-rate and fiber length of source systems are different from those of the target one, we can always gain information related to the target system
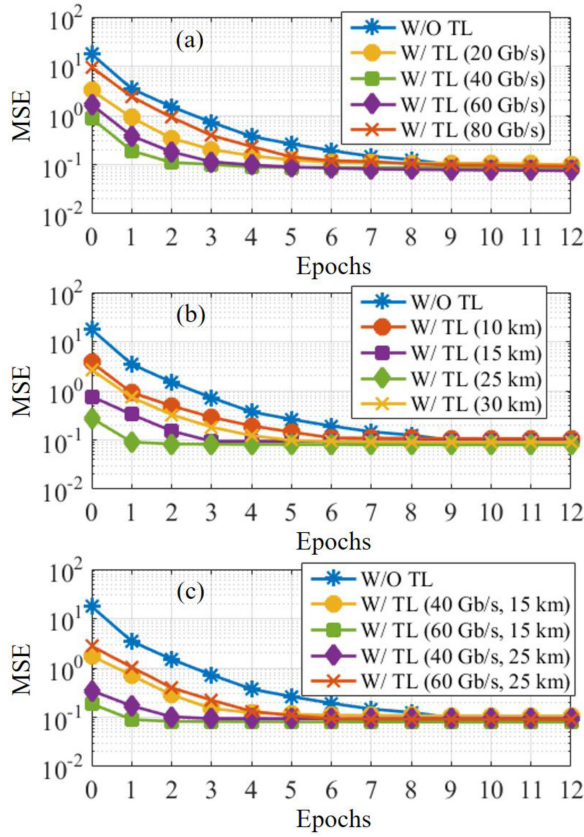
Fig. 3.   MSE versus epochs using TL-aided FNNs from different source systems. TL is applied for source systems with (a) different bit-rates, (b) different fiber lengths, and (c) both effects.

Fig. 4.   BER versus N_{train} using TL-aided FNNs from different source systems. TL is applied for source systems with (a) different bit-rates, (b) different fiber lengths, and (c) both effects.

from the FNNs trained for source systems. Since these NNs have the capability of equalizing the received distorted signals, they "learned" the mixed channel effects which influence the transmitted symbols. This kind of information lies in the parameters of NNs, to be specific, the weights and biases. As such, these source FNNs already preserve a lot of channel information such as CD, frequency chirp, and band-limited effects, where the target system could benefit from by performing TL. As shown in Fig. 3, the epochs to train FNNs can be largely reduced with the help of TL, and all the source systems (even both the bit-rate and fiber length are different) are shown beneficial for the target system. By performing TL using FNNs from the 50-Gb/s 25-km or 60-Gb/s 15-km source system, only 1 epoch is needed, achieving a reduction rate of about 90% compared with the FNN trained without TL. The more similar the source system is to the target one, the better TL performance can be achieved. As shown in Fig. 3(a), in terms of bit-rate, 40-Gb/s source system more resembles 50-Gb/s target system than the 60-Gb/s source system, because the 60-Gb/s signals suffer from power-fading effects after 20-km fiber transmission, which provides additional information the target system may not need. In terms of the impact of fiber length shown in Fig. 3(b), we find that 25-km source system is more similar to the target system compared with the 15-km one. This is because it is easier to delete abundant information from NNs rather than to add new knowledge from
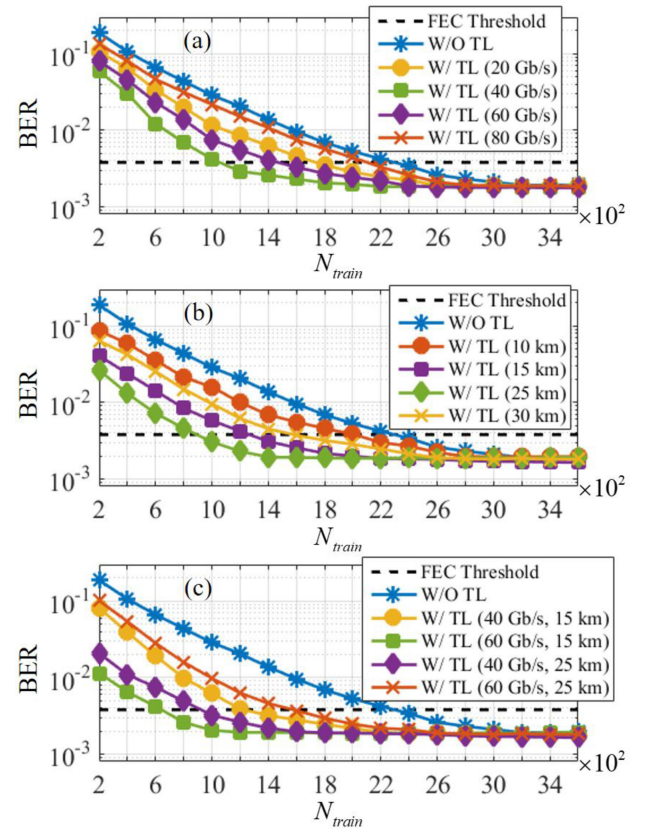
scratch. Among all source systems employed, as depicted in Fig. 3(c), the 60 Gb/s 15-km link is shown to be the most similar one to the target system, with the lowest initial MSE at the start of the training. Even through both the bit-rate and fiber length are changed, it generates the most similar CD effect as the target system since the impact of CD is proportional to the square of bit-rate multiplied by fiber length.

Fig. 4 depicts BER versus the number of training symbols, denoted N_{train}, using TL-aided FNNs from different source systems. The source systems used are the same as shown in Fig. 3, where Fig. 4(a), 4(b), and 4(c) depict different source systems with different bit-rates, different fiber lengths, and both different bit-rates and fiber lengths, respectively. 12 epochs are used to ensure convergence for validation/testing during the TL-aided training process, and the ROP is still at −1 dBm. Roughly similar trends are observed by comparing Fig. 3 and Fig. 4, and the analysis on similarities between the target system and the source ones in Fig. 3 is also applicable for the case in Fig. 4. We can observe that in addition to the reduction of epochs shown in Fig. 3, FNNs trained for all the proposed source systems help significantly reduce the requirment of training symbols as well, and for the best TL case (60-Gb/s 15-km source system), only 1200 training symbols are needed, achieving a reduction rate of 62.5% compared with training without TL. If we only consider the 7% hard-decision forward error correction (FEC) threshold,
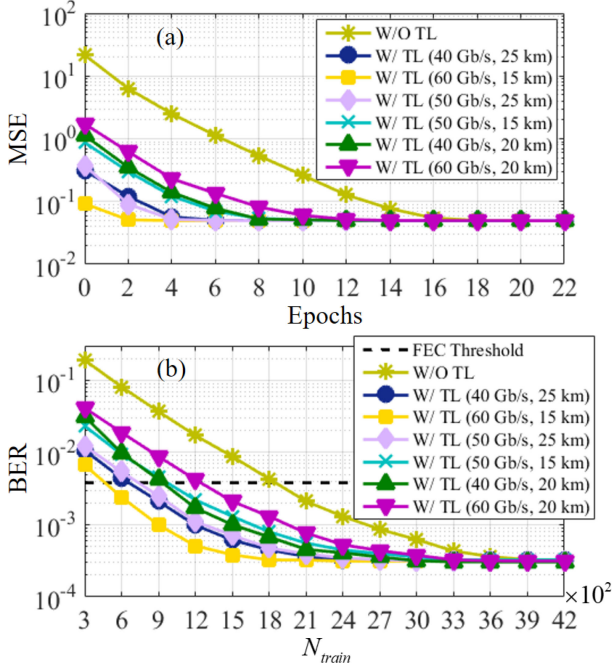
Fig. 5. (a) MSE versus epochs, and (b) BER versus N_{train} when using TL-aided RNNs from different source systems.
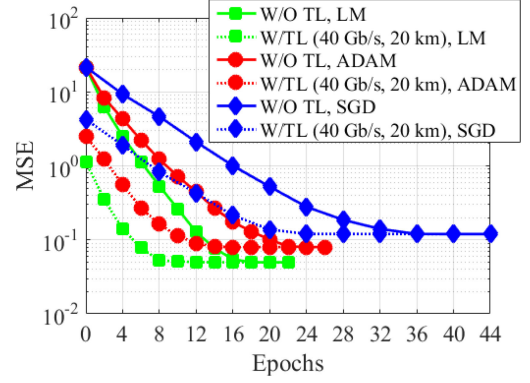

Fig. 6. MSE versus epochs with TL-aided RNNs using different optimizers.
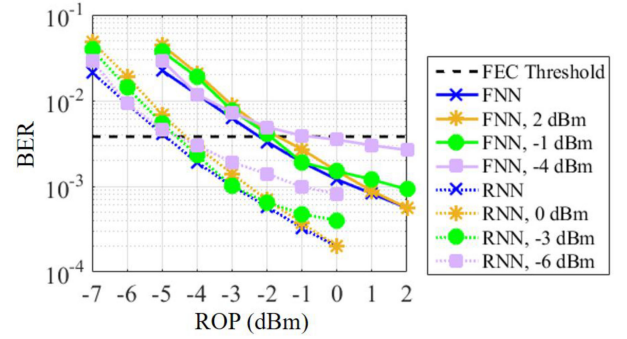

Fig. 7. BER versus ROP using FNNs/RNNs trained at each ROP or at some fixed ROP (FNN: -4, -1, 2 dBm, RNN: -6, -3, 0 dBm).

the requirement for N_{train} could be further relaxed to only about 600.

Besides FNN-based equalizers, RNN-based equalizers are also proved suitable for TL which is shown in Fig. 5, where MSE versus epochs is shown in Fig. 5(a), and BER versus N_{train} is presented in Fig. 5(b) when the ROP is -1 dBm. N_{train} is set at 4500 for Fig. 5(a), while the number of epochs is fixed at 20 in Fig. 5(b). Note that for both cases in Fig. 3 and Fig. 5(a), further increasing the number of training symbols will not affect the performance with or without TL. This is because the training set has already contained enough information of the received symbols. As shown in the figures, RNNs generally requires more epochs and training symbols than FNNs, however, both TL-aided FNNs and TL-aided RNNs reduce the number of epochs and N_{train}, and TL-aided RNNs show similar trends as TL-aided FNNs on the influence of different source systems. For RNN, 60-Gb/s 15-km source system is still the best one for TL among all the source systems employed, and about 87.5% reduction in epochs and 53.8% in training symbols are achieved using RNNs trained for this source system.

Based on the RNN structure, we also investigate on the impact of TL over different optimizers, i.e., LM, ADAM, and SGD, which is depicted in Fig. 6. For simplicity, only the 40-Gb/s 20-km source system is used to show the TL-aided performance. Note that all the parameters of different optimizers are carefully selected to achieve their best results. Among the 3 types of optimizers used, LM shows the best performance in both convergence speed and the achieved MSE for our small-scale regression task, and as such, LM is adopted as the default optimizer in this paper. However, regardless of which optimizer is selected, significant epoch reduction can still be observed with the help of

TL. TL-aided equalizers are not restricted to parameter settings such as the choice of optimizers, because the information of source systems can always contribute to the training process of the target system.

In addition to perturbations such as bit-rate or fiber length, TL is also tested with respect to ROP, and Fig. 7 depicts the BER versus ROP using FNNs or RNNs trained at each ROP or at some fixed ROP values. Note that no retraining process is applied here, and NNs trained at some fixed ROP values are tested for the whole ROP range. As shown in Fig. 7, for both FNNs and RNNs, the NNs trained at a relatively high ROP (E.g., FNN: 2 dBm, RNN: 0 dBm) perform much better than NNs trained at a relatively low ROP (E.g., FNN: -4 dBm, RNN: 6 dBm), because less noise is involved in training when the ROP is high. For FNNs or RNNs trained at low ROP, since these NNs are trained with severe noise, they are not suitable for the environment when the ROP is high, and as such, they may not be able to demonstrate good BER performance especially at high ROP range.

We then investigate on the application of TL considering different target systems (different bit-rates or fiber lengths). Fig. 8(a) illustrates the MSE versus the target bit-rate using FNNs/RNNs trained for 40/50/60-Gb/s 20-km source systems, where same types of NNs are applied for the source and target system. Fig. 8(b) shows MSE versus the target fiber length using FNNs/RNNs trained for 50-Gb/s 15/20/25-km source systems, where only FNNs are used in the target system. The ROP is
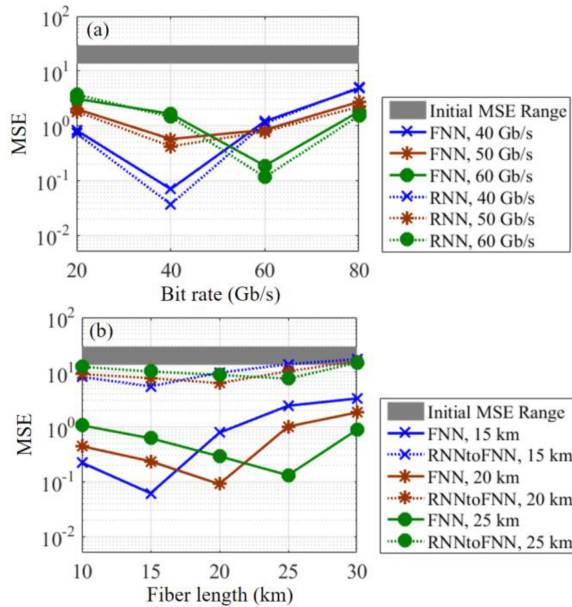
Fig. 8. (a) MSE versus bit-rate (target system: FNN/RNN) using FNNs/RNNs trained for 40/50/60-Gb/s 20-km source systems; (b) MSE versus fiber length (target system: FNN/FNN) using FNNs/RNNs trained for 50-Gb/s 15/20/25-km source systems.

fixed at -1 dBm. Similar to the case in Fig. 7, no retraining step is performed and we only focus on the TL-aided MSE. As shown in Fig. 8(a), for both FNNs and RNNs, the lowest MSE is achieved when the source and target system are selected as the same. When the source system starts to deviate from the target one, the MSE of the target system increases. Compared with the MSE of random initialization (grey part), NNs trained for source systems reduce the MSE by about one order of magnitude, providing a better initial point for training in the target system. We also notice that by setting all the entries of $\mathbf{W}^d$ as 0, FNNs can be well transferred to their corresponding RNNs with the same MSE. However, the reverse step (RNNs transfer to their corresponding FNNs) is not workable as shown in Fig. 8(b), and the MSE values gained from RNNs are found close to the initial MSE range. Since all the weights and biases in RNNs are optimized simultaneously, we cannot expect a good MSE performance by just treating part of them as those of FNNs.

V. Conclusion

In conclusion, TL-aided FNN- and RNN-based equalizers are proposed for nonlinear equalization in a 50-Gb/s 20-km target PAM4 direct detection optical link. The number of training symbols and epochs can be largely shrunk by applying TL to the FNNs or RNNs trained for different types of source systems with varying bit-rates and fiber lengths, which leads to a much shortened NN training time. TL is also demonstrated to be applicable with respect to other factors such as ROP or the conversion from FNNs to RNNs. With the help of TL, the different optical link connections can be established in an expeditious manner, and the fast equalizers are important to support future optical switched data center networks, where the optical interconnects need to be dynamically reconfigured.

References

[1] J. C. Cartledge and A. S. Karar, "100 Gb/s intensity modulation and direct detection," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2809–2814, Aug. 2014.

[2] J. A. Tatum *et al.*, "VCSEL-based interconnects for current and future data centers," *J. Lightw. Technol.*, vol. 33, no. 4, pp. 727–732, Feb. 2015.

[3] A. Ghiasi, "Large data centers interconnect bottlenecks," *Opt. Express*, vol. 23, no. 3, pp. 2085–2090, Jan. 2015.

[4] X. Pang *et al.*, "200 Gbps/lane IM/DD technologies for short reach optical interconnects," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 492–503, Dec. 2019.

[5] F. Li *et al.*, "Optical amplifier-free 100 Gbit/s/lamda PAM-N transmission and reception in O-band over 40-km SMF with 10-G class DML," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2019, Paper Tu2F.4.

[6] C. Yang *et al.*, "IM/DD-based 112-Gb/s/lambda PAM-4 transmission using 18-Gbps DML," *IEEE Photon. J.*, vol. 8, no. 3, pp. 1–7, Jun. 2016.

[7] K. Zhong *et al.*, "Experimental study of PAM-4, CAP-16, and DMT for 100 Gb/s short reach optical transmission systems," *Opt. Express*, vol. 23, no. 2, pp. 1176–1189, Jan. 2015.

[8] K. Zhong, X. Zhou, J. Huo, C. Yu, C. Lu, and A. P. T. Lau, "Digital signal processing for short-reach optical communications: A review of current technologies and future trends," *J. Lightw. Technol.*, vol. 36, no. 2, pp. 377–400, Jan. 2018.

[9] L. Yi *et al.*, "Neural network-based equalization in high-speed PONs," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2020, Paper T4D.3.

[10] A. G. Reza and J. K. K. Rhee, "Nonlinear equalizer based on neural networks for PAM-4 signal transmission using DML," *IEEE Photon. Technol. Lett.*, vol. 30, no. 15, pp. 1416–1419, Aug. 2018.

[11] J. Estaran *et al.*, "Artificial neural networks for linear and non-linear impairment mitigation in high-baudrate IMDD systems," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2016, Paper M.2.B.2.

[12] C. Y. Chuang *et al.*, "Employing deep neural network for high speed 4-PAM optical interconnect," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2017, Paper W.2.D.2.

[13] Z. Xu *et al.*, "Cascade recurrent neural network enabled 100-Gb/s PAM4 short-reach optical link based on DML," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2020, Paper W2A.45.

[14] Z. Xu *et al.*, "Cascade recurrent neural network-assisted nonlinear equalization for a 100 Gb/s PAM4 short-reach direct detection system," *Opt. Lett.*, vol. 45, no. 15, pp. 4216–4219, Aug. 2020.

[15] C. Y. Chuang *et al.*, "Convolutional neural network based nonlinear classifier for 112-Gbps high speed optical link," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2018, Paper W2A.43.

[16] C. Ye *et al.*, "Recurrent neural network (RNN) based end-to-end nonlinear management for symmetrical 50Gbps NRZ PON with 29dB+ loss budget," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2018, Paper Mo4B.3.

[17] Z. Yang *et al.*, "Radial basis function neural network enabled C-band 4 × 50 Gb/s PAM-4 transmission over 80 km SSMF," *Opt. Lett.*, vol. 43, no. 15, pp. 3542–3545, Aug. 2018.

[18] B. Karanov *et al.*, "End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks," *Opt. Express*, vol. 27, no. 14, pp. 19650–19663, 2019.

[19] B. Karanov *et al.*, "Experimental investigation of deep learning for digital signal processing in short reach optical fiber communications," in *Proc. IEEE Workshop Signal Process. Syst. (SiPS)*, Oct. 2020, pp. 1–6. [Online]. Available: https://arxiv.org/pdf/2005.08790.pdf

[20] Z. Xu *et al.*, "Computational complexity comparison of feedforward/radial basis function/recurrent neural network-based equalizer for a 50-Gb/s PAM4 direct-detection optical link," *Opt. Express*, vol. 27, no. 25, pp. 36953–36964, Dec. 2019.

[21] B. Karanov *et al.*, "End-to-end deep learning of optical fiber communications," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4843–4855, 2018.

[22] Z. Xu *et al.*, "Transfer learning aided neural networks for nonlinear equalization in short-reach direct detection systems," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2020, Paper T4D.4.

[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.

[24] D. Azzimonti *et al.*, "Active vs transfer learning approaches for QoT estimation with small training datasets," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2020, Paper M4E.1.

[25] L. Xia *et al.*, "Transfer learning assisted deep neural network for OSNR estimation," *Opt. Express*, vol. 27, no. 14, pp. 19398–19406, Jun. 2019.

[26] D. Wang *et al.*, "Comprehensive eye diagram analysis: A transfer learning approach," *IEEE Photon. J.*, vol. 11, no. 6, pp. 1–19, Dec. 2019.

[27] J. Zang *et al.*, "Fast remodeling for nonlinear distortion mitigation based on transfer learning," *Opt. Lett.*, vol. 44, no. 17, pp. 4243–4246, Sep. 2019.