Enhanced Recurrent Neural Network Equalization based on Hidden Feature Extraction Learning for Optical Interconnect

Chuanchuan Yang[1,*], Yunfeng Gao[1,2], Jiaxing Wang[3], Hongbin Li[1], Constance J. Chang-Hasnain[3]
[1] School of Electronics, Peking University, Beijing 100871, China
[2]CITIC Group Corporation, Beijing 100020, China
[3]Shenzhen Berxel Photonics Co. Ltd., Shenzhen 518055, China
[*]E-mail address: yangchuanchuan@pku.edu.cn

Abstract: We propose a hidden feature extraction learning method for RNN equalization to improve training efficiency without increasing computational burden. Superior BER is demonstrated in 288 Gb/s 100 m VCSEL-MMF interconnect compared with black-box training strategy. © 2024 The Author(s)

1. Introduction

In recent years, the development of bandwidth-hungry applications has brought an exponential growth in data center traffic. In order to counteract the resulting bandwidth limitations and nonlinear distortions that affect VCSEL-based intensity modulation and direct detection (IM/DD) interconnects at high data rates, advanced digital signal processing (DSP) techniques have been extensively investigated [1]. Particularly, neural networks (NNs) provide an attractive alternative to traditional model-based approaches to overcome complex and challenging characteristic modeling [2,3]. Thanks to the powerful capability in dealing with high complexity and high dimensionality problems, NNs are able to achieve excellent performance with complex structures [4].

Meanwhile, deep learning (DL) techniques, which can approximate any nonlinear function, enable the design of communication systems by carrying out the optimization in a single end-to-end process including the transceivers as well as communication channels [5]. In contrast to traditional block-based signal processing methods, it achieves global optimum by directly fitting from raw data to the final target. Actually, a deep neural network with multiple layers autonomously acquires a robust representation of the input message through numerous iterations. However, the training process belongs to an entire black-box strategy. In order to ensure the accuracy and effectiveness of training results, a large amount of data input is highly required [6].

In this paper, an enhanced deep recurrent neural network (RNN) equalization method based on hidden feature extraction (HFE) learning is proposed, which owns more effective training talent. The validity and reliability of the proposed method are experimentally demonstrated in several VCSEL-based high-speed interconnect scenarios up to 288 Gb/s over 100 m link. Moreover, superior bit error ratio (BER) performance is achieved compared with the traditional black-box training strategy.

2. Enhanced RNN Equalizer via Hidden Feature Extraction Learning

As a consequence of nonlinear distortions induced by transceivers and the fiber, the short-reach optical interconnect behaves as a non-Gaussian channel. Moreover, arising from the interaction of inter-symbol interference (ISI) and square-law detection, the channel has nonlinear memory effects [7]. For neural networks with linearly-connected neurons, the nonlinearity is predominantly fitted by the activation function with an arduous burden in training.

Fig. 1 shows the complete flowchart of the proposed enhanced HFE learning method. We introduce the hidden

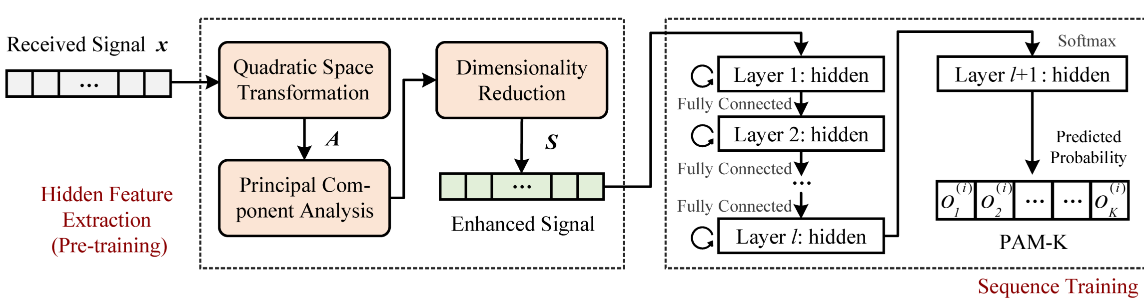

Fig. 1. Flowchart of the proposed HFE learning method, including the feature (pre-training) and sequence learning process.

feature extraction as pre-training before the sequence training. It should be noted that hidden features refer to characteristics that cannot be observed directly from the raw data but are critical for improving NN-based signal detection. It can be extracted through feature learning process as follows. Based on the original received signal sequence $\{x(t)\}$, quadratic space transformation characterizes the nonlinear memory effects by considering the repeatable product of a certain length of received signal, which is expressed as

$$[c_1^{(t)},...,c_{L+1}^{(t)},...,c_{2L+1}^{(t)},c_{2L+2}^{(t)},...,c_p^{(t)},...,c_{N_f}^{(t)}]=[x(t+L),...,x(t),...,x(t-L),x^2(t+L),...,x(t-r_1)x(t-r_2),...,x^2(t-L)]. \quad (1)$$

where $-L \le r_1 \le r_2 \le L$ and L is the number of adjacent symbols required for transformation. The number of characterized features, denoted as N_f, is equal to $(2L+1)(L+2)$. The neural network can learn directly with the feature component $c_p^{(t)}(p=1,...,N_f)$ as input, but its size will be further expanded compared to the raw data. Consequently, principal component analysis is introduced for further compression.

Define the pre-compression feature matrix $\boldsymbol{A}$:

$$\boldsymbol{A}=[\boldsymbol{C}_1,...,\boldsymbol{C}_p,...,\boldsymbol{C}_{N_f}]_{N_t\times N_f},\ \boldsymbol{C}_p=[c_p^{(t)},...,c_p^{(t+N_t-1)}]^T. \quad (2)$$

N_t denotes the length of training data. Due to the fact that $\boldsymbol{A}^T\boldsymbol{A}$ is a real symmetric matrix and has an equal rank with $\boldsymbol{A}$, there must exist a N_f-order orthogonal matrix $\boldsymbol{V}=(V_{ij})_{N_f\times N_f}$ that satisfies

$$\boldsymbol{V}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{V}=\mathrm{diag}\left(\lambda_1,...,\lambda_m,0,...,0\right)=\begin{bmatrix}\boldsymbol{\mathcal{D}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}^2. \quad (3)$$

where m is the rank of $\boldsymbol{A}$ and λ_g (g=1,2,...m) is the non-zero eigenvalue of $\boldsymbol{A}$. $\boldsymbol{\mathcal{D}}$ is the diagonal matrix composed of singular values as $\sqrt{\lambda_g}$. As it is feasible to keep $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_m$, the features can be characterized by the former N_{rd}-dimensional sub-blocks ($1 \le N_{rd} \le N_f$). The hidden features after dimensionality reduction can be derived as

$$\boldsymbol{S}_{N_t\times N_{rd}}=\boldsymbol{A}_{N_t\times N_f}V_{N_f\times N_{rd}}. \quad (4)$$

The sequence learning is performed based on RNN. Through connection feedback and parameter sharing, it allows the current output to influence the subsequent input in the same node, exhibiting time-dependent dynamic tracking capability. The depth fitting in the input/output direction is also achieved by increasing the number of hidden layers. The forward propagation process of RNN with l hidden layers can be expressed as

$$\boldsymbol{h}_j^{(t)}=\begin{cases}\sigma(\boldsymbol{W}_{r,j}\boldsymbol{h}_j^{(t-1)}+\boldsymbol{W}_{s,j}\boldsymbol{s}^{(t)}+\boldsymbol{b}_j),\ j=1\\ \sigma(\boldsymbol{W}_{r,j}\boldsymbol{h}_j^{(t-1)}+\boldsymbol{W}_{s,j}\boldsymbol{h}_{j-1}^{(t)}+\boldsymbol{b}_j),\ j\in(1,l]\end{cases},\ \boldsymbol{o}^{(t)}=\mathrm{softmax}\ (\boldsymbol{W}_{s,l+1}\boldsymbol{h}_l^{(t)}+\boldsymbol{b}_{l+1}). \quad (5)$$

where σ(.) is the activation function. $\boldsymbol{W}_{s,j}$, $\boldsymbol{W}_{r,j}$, $\boldsymbol{b}_j$ are the weights of neurons in the j-th layer and $\boldsymbol{h}_j^{(t)}$ is the corresponding hidden layer state. All parameters are learned by the back propagation through time algorithm.

After the training stage described above, the proposed RNN equalizer works online. We propose a new HFE learning method and after the training, RNN also works in the classical way. For the PAM-K signal, the predicted probability of each modulation amplitude $o_k^{(t)}(k=1,2,\cdots,K)$ can be obtained by softmax function. Based on the maximum posterior probability rule, the signal detection results after equalization are derived as

$$y^{(t)}=arg\max_k\left\{o_k^{(t)}\right\},\ \sum_{k=1}^{K}o_k^{(t)}=1. \quad (6)$$

Generally, N_{rd}=1 is sufficient to meet the performance requirements. Compared to the traditional training method for RNN, the input dimension remains the same after HFE, which means we try to improve the training efficiency of neural networks without increasing computational burden of training process.

3. Experimental Results

The experimental setup and digital signal processing (DSP) diagram are depicted in Fig. 2(a). At the transmitter side, PAM-8 symbols are generated by an arbitrary waveform generator (AWG) with 120 GSa/s sampling rate. We employ 850 nm VCSEL from Berxel as light source and obtain a 3 dB bandwidth of 23 GHz with an optimized 8 mA bias current. The operating bias is set with a 65 GHz bias-T and the optical signal is coupled into the 100 m OM-4 MMF via a lensed fiber. At the receiver side, the optical signal is directly detected by a photoreceiver with 22 GHz bandwidth. To reduce the interference of oscilloscope noise floor, the received signal is amplified by a 22 dB electrical amplifier (EA) before being sampled. The real-time digital storage oscilloscope (DSO) captures the signal at 160 GSa/s rate. The frequency response of the system is measured via the PNA-network analyzer (Keysight N5247A) and the result is depicted in Fig. 2(b).

While evaluating the performance of RNN trained by the proposed HFE method, RNN trained in the traditional way is adopted as the benchmark. The number of hidden layers l is kept to 3, and rectified linear unit (ReLU) is used

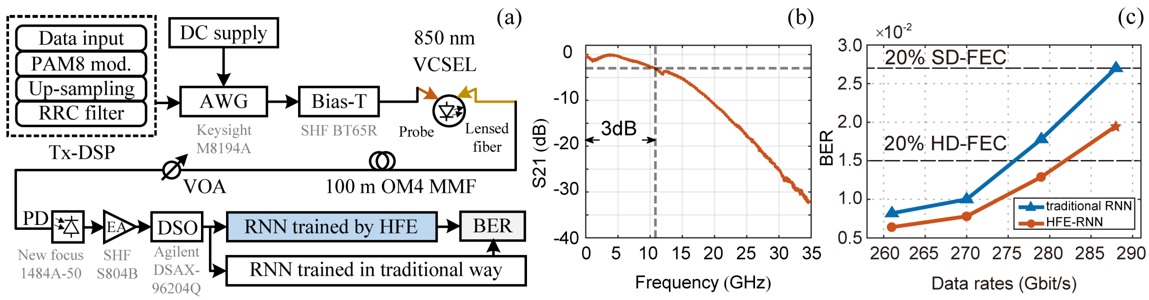

Fig. 2. (a). Experimental configuration of a PAM-8 signal transmission system utilizing 850 nm VCSEL. (b) Measured frequency response of the 100 m VCSEL-MMF link. (c) Measured BER performance of RNN trained in the traditional way and RNN trained by the proposed HFE method at different transmission rates.

as the activation function σ(.) to avoid the vanishing gradients in training. A dataset containing 510,0000 PAM-8 symbols are obtained following [8]. We have extracted 450,0000 of them to construct the training set, while the rest are put into the test set.

Fig. 2(c) shows the BER performance at different transmission rates. Compared to the traditional RNN, RNN trained by HFE obtains a growing BER gain as the data rates increase. This is attributed to the pre-training process that extracts hidden features and guides the deep RNN in channel learning, where L and N_{rd} are set to 2 and 1 respectively for optimal performance. The experimental results present that BER of 288 Gb/s PAM-8 signal transmission based on HFE-RNN over 100 m OM4 MMF is much lower than 20% soft-decision forward error correction (FEC) threshold.

Fig. 3 compares the performance of the two training methods for RNN. As epochs gradually increase, the loss shows a trend of rapid reduction followed by a slow decrease until convergence. It should be noted that slight fluctuation of the loss during the iteration process is normal, which does not affect the achievement of convergence. This indicates that the training process effectively learns the channel characteristics of the high-capacity VCSEL-MMF system. Compared with the traditional method, HFE has lower loss performance as well as higher efficiency in the same number of epochs.

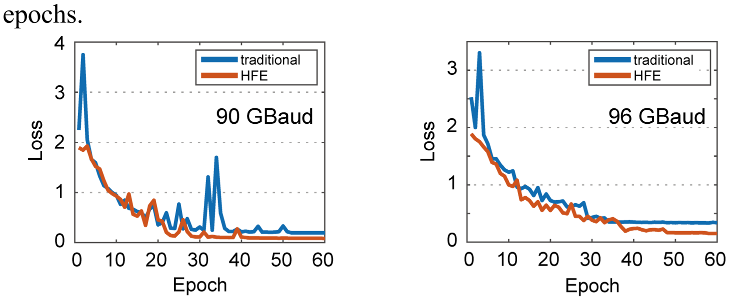

Fig. 3. Performance comparison of different training methods for RNN.

4. Conclusion

An enhanced RNN method based on HFE learning is proposed to realize effective signal detection. Compared with the traditional black-box structure of end-to-end learning, HFE obtains enhanced feature input through quadratic space transformation, principal component analysis and dimensionality reduction. It is capable of improving the learning efficiency of NNs without increasing the computational burden. Using the new method, single-lane 288 Gb/s signal transmission over 100 m MMF is implemented with BER well below the 20% SD-FEC threshold.

5. Acknowledgement

This work was supported by National Key R&D Program of China (No. 2019YFB1802904), the Shenzhen Science and Technology Program (No. KQTD20200820113053102, No. ZDSYS20220325163600001) and the Guangdong Major Talent Introduction Project (No. 2021ZT09X328).

6. References

[1] M. Srinivasan *et al.*, *JLT*, vol. 41, no. 11, pp. 3261-3277 (2023).
[2] K. Zhong *et al.*, *JLT*, vol. 36, no. 2, pp. 377-400 (2018).
[3] M. Schädler *et al.*, *ECOC*, Th1C-1 (2022).
[4] L. Huang *et al.*, *JLT*, vol. 40, no. 14, pp. 4528-4538 (2022).
[5] C. Häger, *ECOC*, Tutorial SC4 (2022).
[6] Z. Xu *et al.*, *IEEE PTL*, vol. 33, no. 9, pp. 425-428 (2021).
[7] Y. Gao *et al.*, *JLT*, vol. 40, no. 12, pp. 3622-3632 (2022).
[8] T.A. Eriksson *et al.*, *IEEE PTL*, vol. 29, no. 23, pp. 2091-2094 (2017)