Self-Supervised Learning for Neural-Network-Based Perturbative Fiber Nonlinearity Compensation

Du Tang(1), Yingjie Jiang(1), Zhen Wu(1), and Yaojun Qiao(1)

(1) The State Key Laboratory of Information Photonics and Optical Communications, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China, qiao@bupt.edu.cn

Abstract *A self-supervised learning scheme is proposed for neural-network-based perturbative fiber nonlinearity compensation with a designed proxy task by using phase-conjugated unknown sequences. The self-supervised trained model can be applied directly to the corresponding dual-polarization system with no obvious performance penalty compared with the supervised one.* ©2023 The Author(s)

Introduction

Fiber nonlinearity has become a major limiting factor for further improving the capacity of long-haul coherent optical transmission systems[1]. To overcome this problem, several machine-learning approaches have been introduced into fiber nonlinearity compensation (NLC) through digital signal processing (DSP) over the last few years[2]–[13]. On this ground, with perturbation triplets as the input features, neural-network-based perturbative NLC (NN-PNLC) algorithms have been intensively studied recently[2]–[6]. This kind of NLC algorithm can be trained without prior knowledge of the fiber link parameters and can be re-trained according to the change of link based on a newly collected dataset[3], becoming a flexible and effective NLC solution.

However, the current NN-PNLC adopts a supervised training scheme, which requires huge training datasets (sometimes up to 2^{20}[5]) to prevent over-fitting and requires accurate knowledge of the sending data as a reference. Although data augmentation (DA) can be used if insufficient datasets are collected[3], it still needs to know the exact knowledge of these insufficient datasets and the DA itself adds the processing complexity in the training process. In addition, accurately knowing a large amount of reference data is not easy. Without storing these data in advance, the usual practice is to generate the reference data according to the same rules as the transmitter, which still consumes hardware resources and may fall into possible pitfalls[14]. Therefore, designing a training scheme for NN-PNLC without knowing the transmitted data is highly desirable for practical implementation.

In this work, to the best of our knowledge, we propose the first self-supervised learning (SSL) scheme for NN-PNLC. The proposed SSL scheme is realized by designing a proxy task[15] to automatically extract the target perturbation terms from phase-conjugated unknown sequences. When the SSL is completed, the trained model can be applied directly to the corresponding dual-polarization (DP) system. The effectiveness of the proposed SSL scheme is verified numerically in a 32-GBaud DP-16QAM 2800-km coherent system and the SSL model exhibits no obvious performance penalty compared with the supervised model.

Self-Supervised Learning for NN-PNLC

According to the first-order perturbation theory, with large chromatic dispersion assumption, the intra-channel fiber nonlinear perturbation terms for the k-th DP symbol can be approximated as[16]

$$N^x_{k,NL} = \sum_{m,n} P_0^{3/2} \left(X_{n+k} X^*_{m+n+k} X_{m+k} + Y_{n+k} Y^*_{m+n+k} X_{m+k} \right) C_{m,n} \tag{1}$$

$$N^y_{k,NL} = \sum_{m,n} P_0^{3/2} \left(Y_{n+k} Y^*_{m+n+k} Y_{m+k} + X_{n+k} X^*_{m+n+k} Y_{m+k} \right) C_{m,n} \tag{2}$$

where P_0, X_k and Y_k, and $C_{m,n}$ denote the launch power, transmitted symbol sequences for the x- and y-polarization, and nonlinear perturbation coefficients, respectively. m and n are symbol indices with respect to the k-th symbol. The received symbols can be denoted as $\hat{X}_m$ and $\hat{Y}_m$. With perturbation triplets $\hat{X}_{n+k}\hat{X}^*_{m+n+k}\hat{X}_{m+k} + \hat{Y}_{n+k}\hat{Y}^*_{m+n+k}\hat{Y}_{m+k}$ as the input features, NN-PNLC outputs the estimated perturbation terms $\hat{N}^{x/y}_{k,NL}$[2] and can be trained by a supervised learning scheme: by minimizing the mean square error (MSE) between the transmitted symbols and the received symbols after eliminating the estimated perturbation terms. Take x-polarization as an example, assuming that the linear impairments have been well compensated, the equation of MSE can be written and further derived as

$$\mathrm{MSE} = \frac{1}{B}\sum_{k=1}^{B}\left|\left(\hat{X}_k - \hat{N}_{k,NL}^{x}\right) - X_k\right|^2 \tag{3a}$$

$$= \frac{1}{B}\sum_{k=1}^{B}\left|\left(\hat{X}_k - X_k\right) - \hat{N}_{k,NL}^{x}\right|^2 \tag{3b}$$

$$= \frac{1}{B}\sum_{k=1}^{B}\left|N_{k,ASE}^{x} + N_{k,NL}^{x} - \hat{N}_{k,NL}^{x}\right|^2 \tag{3c}$$

where B denotes the batch size and $N_{k,ASE}^{x}$ denotes the amplified spontaneous emission (ASE) noise. Through the derivation of Eq. (3b) to (3c) of Eq. (3), it can be found that the MSE of NN-PNLC mainly depends on the difference between the reference nonlinear perturbation terms $N_{k,NL}^{x}$ and the estimated nonlinear perturbation terms $\hat{N}_{k,NL}^{x}$. If a certain technique can be used to extract the reference$N_{k,NL}^{x}$ from the received data, then the restriction that the transmitted data must be known will be lifted.

Fortunately, the phase-conjugated twin waves (PCTW)[17] technique can help to achieve this goal. For a long time, by sacrificing one polarization ($Y_k = X_k^*$), the PCTW technique is often used for fiber nonlinear compensation with simple operation $\hat{X}_k + \hat{Y}_k^*$ based on the relation $N_{k,NL}^{y} = -(N_{k,NL}^{x})^*$. However, if we look at PCTW from a completely opposite perspective, it can be used as a perturbation extractor by:

$$N_{k,ext}^{x} = \frac{1}{2}\left(\hat{X}_k - \hat{Y}_k^*\right) = N_{k,ASE} + N_{k,NL}^{x} \tag{4}$$

where $N_{k,ASE} = \frac{1}{2}(N_{k,ASE}^{x} - N_{k,ASE}^{y^*})$ denotes the ASE noise after the perturbation extractor. The results of Eq. (4) are exactly the first two terms needed in the square calculation of Eq. (3c) so that the calculation of MSE only needs the received unknown data $\hat{X}_k$ and $\hat{Y}_k$ now and no longer depends on the known transmitted data.

Then, if phase-conjugated sequences are temporarily used for the training phase of NN-PNLC in a DP system, the perturbation terms can be extracted just from the received unknown data, updating the supervised learning (SL) to self-supervised learning (SSL). The diagram of the proposed SSL scheme is illustrated in Fig. 1. In addition to the triplets calculation, a proxy task is carried out to extract the reference perturbation

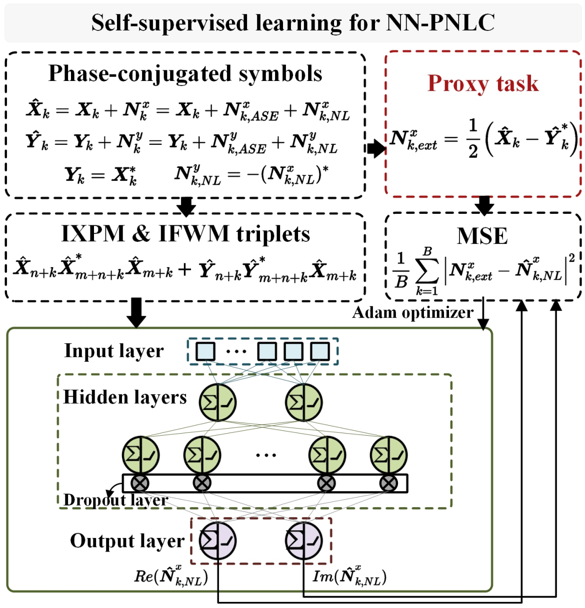

Fig. 1: The diagram of the proposed SSL for NN-PNLC.

terms $N_{k,ext}^{x}$ from $\hat{X}_k$ and $\hat{Y}_k$ according to Eq. (4). The NN structure is similar to the one in[2] with an input layer, 2 dense layers with 2 and 10 neurons followed by Leaky ReLU activation function, respectively, and one dense layer outputs the real and imaginary parts of the estimated perturbation terms $\hat{N}_{k,NL}^{x}$. Finally, MSE is calculated with $\frac{1}{B}\sum_{k=1}^{B}\left|N_{k,ext}^{x} - \hat{N}_{k,NL}^{x}\right|^2$. Finally, the NN is optimized by Adam with a learning rate of 1×10^{-4} and batch size of B = 1024.

To evaluate the performance of NN-PNLC with the proposed SSL scheme (SSL-NN-PNLC) and with the conventional SL scheme (SL-NN-PNLC), as shown in Fig. 2 (a), we simulated a 32-GBaud DP-16QAM coherent system. Note that at the training stage, the transmitted symbols of the two polarizations are phase-conjugated but unknown at the receiver while at the implementation stage, the transmitted symbols of the two polarizations return to independent. The signal is pulse shaped by a root-raised cosine (RRC) filter with a 0.01 roll-off factor and 50% pre-chromatic dispersion compensation (CDC) is performed at the transmitter DSP processing. The laser linewidth is 100 kHz and no other transceiver impairments are considered. The signal is then transmitted through 35 spans of 80-km standard single-mode fiber (SSMF) with an attenuation coefficient of 0.2 dB/km, a dispersion parameter of 17 ps/nm/km, and a nonlinear parameter of 1.3 $W^{-1}km^{-1}$. Every span is followed by an erbium-doped-fiber amplifier (EDFA) with a 4.5-dB noise figure. After the coherent receiver, the obtained signal will be processed through DSP. The DSP process of

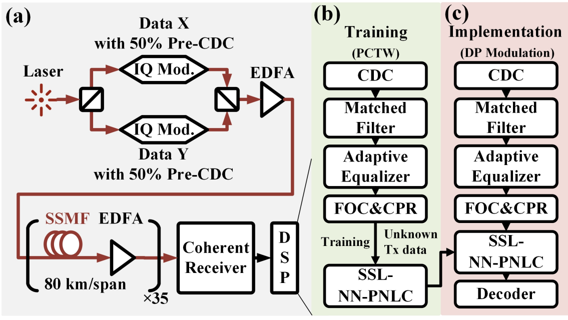

Fig. 2: The diagram of (a) the studied numerical system, (b) DSP process of the training stage of SSL-NN-PNLC, and (c) DSP process of the implementation stage of SSL-NN-PNLC.

the training and implementation stage of SSL-NN-PNLC are shown in Fig. 2 (b) and (c), respectively. Both two stages contain a 50% post-CDC, a matched filter, an adaptive equalizer, and a module of frequency offset compensation (FOC) and carrier phase recovery (CPR). At the training stage, SSL is enabled with the help of phase-conjugated unknown sequences with a size of 10^5. After SSL is completed, the model is fixed and can be implemented after CPR with the DP modulation format at the implementation stage, and 10^5 symbols are used to evaluate the performance with $Q^2 = 20\log_{10}[\sqrt{2}\,\mathrm{erfc}^{-1}(2\,\mathrm{BER})]$ where erfc^{-1} denotes the inverse complementary error function and BER denotes the bit error rate.

Results and Discussion

First, the MSE evolutions of SL-NN-PNLC and SSL-NN-PNLC both with 2097 triplets are shown in Fig. 3 (a). Both SSL-NN-PNLC and SL-NN-PNLC tend to converge after about 100 epochs. Moreover, the absolute MSE value of SSL-NN-PNLC after convergence is lower than the one of SL-NN-PNLC. This is reasonable because the proxy task also halves the variance of ASE noise. Since $N^x_{k,ext}$ and $N^x_{k,ASE} + N^x_{k,NL}$ are no longer the same, the comparison of the absolute value of the MSE after convergence cannot indicate the performance of the two models.

Next, Fig. 3 (b) illustrates the Q^2 performance of CDC, SL-NN-PNLC, and SSL-NN-PNLC evaluated with DP signals. The triplets input size of SL-NN-PNLC and SSL-NN-PNLC are also 2097. SSL-NN-PNLC provides about 1-dB Q^2 improvement compared with CDC and the optimal launch power is improved from -1 dBm to 1 dBm, similar to the performance of SL-NN-PNLC with only a 0.1-dB Q^2 penalty. This penalty is caused by the imperfection of the automatically extracted reference perturbation terms by the proxy task and it is not obvious compared with the improvement.

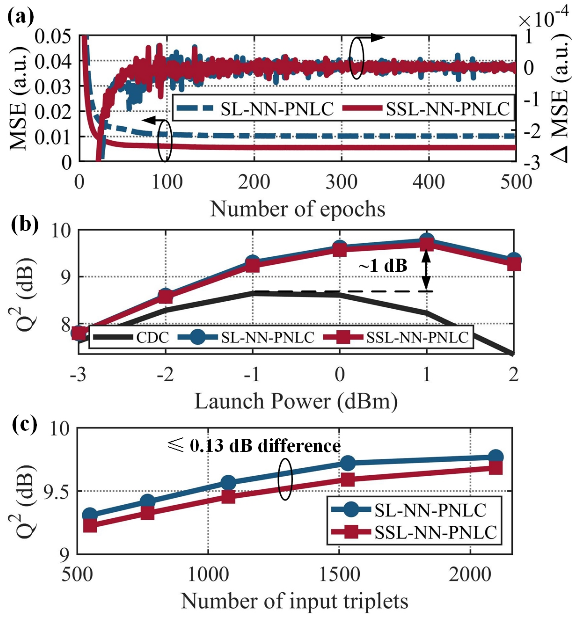

Fig. 3: (a) The MSE evolution of SL and the proposed SSL. (b) Performance of CDC, SL-NN-PNLC, and SSL-NN-PNLC as a function of launch power. (c) Comparison between the best performance as a function of the number of input triplets.

Finally, we also made a comparison between the best Q^2 performance of SL-NN-PNLC and SSL-NN-PNLC as a function of the number of input triplets. The results in Fig. 3 (c) indicate that similar to SL-NN-PNLC, the performance of SSL-NN-PNLC also increases with the number of input triplets, just within 0.13-dB penalty compared with SL-NN-PNLC. Above all, it is verified that SSL-NN-PNLC is still effective in DP systems with phase-conjugated unknown symbols in the training stage. The proposed SSL may also be extended to other learned versions of PNLC.

Conclusions

We have proposed an SSL scheme for NN-PNLC. By temporarily changing the signal of the two polarizations phase-conjugated at the training stage, the reference perturbation terms can be automatically extracted through a proxy task without the knowledge of transmitted symbols. With only a slight performance penalty compared to SL, the proposed SSL scheme removes the restriction that the transmitted data must be known during the training of NN-PNLC.

Acknowledgements

This work is supported in part by National Natural Science Foundation of China under Grant 62271080, in part by Fund of State Key Laboratory of IPOC (BUPT) under Grant IPOC2022ZT06, and in part by BUPT Excellent Ph.D. Students Foundation under Grant CX2022102.

References

[1] O. Vassilieva, I. Kim, and T. Ikeuchi, "Enabling technologies for fiber nonlinearity mitigation in high capacity transmission systems", *Journal of Lightwave Technology*, vol. 37, no. 1, pp. 50–60, 2019. DOI: 10.1109/JLT.2018.2877310.

[2] S. Zhang, F. Yaman, K. Nakamura, T. Inoue, V. Kamalov, L. Jovanovski, V. Vusirikala, E. Mateo, Y. Inada, and T. Wang, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks", *Nature communications*, vol. 10, no. 1, p. 3033, 2019. DOI: 10.1038/s41467-019-10911-9.

[3] V. Neskorniuk, P. J. Freire, A. Napoli, B. Spinnler, W. Schairer, J. E. Prilepsky, N. Costa, and S. K. Turitsyn, "Simplifying the supervised learning of kerr nonlinearity compensation algorithms by data augmentation", in *2020 European Conference on Optical Communications (ECOC)*, Brussels, Belgium, 2020, pp. 1–4. DOI: 10.1109/ECOC48923.2020.9333417.

[4] S. Luo, S. K. Orappanpara Soman, L. Lampe, J. Mitra, and C. Li, "Learning for perturbation-based fiber nonlinearity compensation", in *2022 European Conference on Optical Communication (ECOC)*, Basel, Switzerland, 2022, pp. 1–4. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=ECEOC-2022-We1C.4.

[5] P. J. Freire, Y. Osadchuk, B. Spinnler, W. Schairer, A. Napoli, N. Costa, J. E. Prilepsky, and S. K. Turitsyn, "Experimental study of deep neural network equalizers performance in optical links", in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, San Francisco, USA, 2021, pp. 1–3. DOI: 10.1364/OFC.2021.M3H.2.

[6] Y. Gao, Z. A. El-Sahn, A. Awadalla, D. Yao, H. Sun, P. Mertz, and K.-T. Wu, "Reduced complexity nonlinearity compensation via principal component analysis and deep neural networks", in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, USA, 2019, pp. 1–3. DOI: 10.1364/OFC.2019.Th2A.49.

[7] S. Deligiannidis, C. Mesaritakis, and A. Bogris, "Performance and complexity evaluation of recurrent neural network models for fibre nonlinear equalization in digital coherent systems", in *2020 European Conference on Optical Communications (ECOC)*, Brussels, Belgium, 2020, pp. 1–4. DOI: 10.1109/ECOC48923.2020.9333288.

[8] C. Häger and H. D. Pfister, "Physics-based deep learning for fiber-optic communication systems", *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 280–294, 2021. DOI: 10.1109/JSAC.2020.3036950.

[9] Q. Fan, C. Lu, and A. P. T. Lau, "Combined neural network and adaptive dsp training for long-haul optical communications", *Journal of Lightwave Technology*, vol. 39, no. 22, pp. 7083–7091, 2021. DOI: 10.1109/JLT.2021.3111437.

[10] J. Ding, T. Liu, T. Xu, W. Hu, S. Popov, M. S. Leeson, J. Zhao, and T. Xu, "Intra-channel nonlinearity mitigation in optical fiber transmission systems using perturbation-based neural network", *Journal of Lightwave Technology*, vol. 40, no. 21, pp. 7106–7116, 2022. DOI: 10.1109/JLT.2022.3200827.

[11] H. Dzieciol, T. Koike-Akino, Y. Wang, and K. Parsons, "Inverse regular perturbation with ml-assisted phasor correction for fiber nonlinearity compensation", *Optics Letters*, vol. 47, no. 14, pp. 3471–3474, 2022. DOI: 10.1364/OL.460929.

[12] Y. Liu, V. Sanchez, P. J. Freire, J. E. Prilepsky, M. J. Koshkouei, and M. D. Higgins, "Attention-aided partial bidirectional rnn-based nonlinear equalizer in coherent optical systems", *Optics Express*, vol. 30, no. 18, pp. 32 908–32 923, 2022. DOI: 10.1364/OE.464159.

[13] P. J. Freire, A. Napoli, D. A. Ron, B. Spinnler, M. Anderson, W. Schairer, T. Bex, N. Costa, S. K. Turitsyn, and J. E. Prilepsky, "Reducing computational complexity of neural networks in optical channel equalization: From concepts to implementation", *Journal of Lightwave Technology*, pp. 1–26, 2023. DOI: 10.1109/JLT.2023.3234327.

[14] T. A. Eriksson, H. Bülow, and A. Leven, "Applying neural networks in optical communication systems: Possible pitfalls", *IEEE Photonics Technology Letters*, vol. 29, no. 23, pp. 2091–2094, 2017. DOI: 10.1109/LPT.2017.2755663.

[15] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 4037–4058, 2021. DOI: 10.1109/TPAMI.2020.2992393.

[16] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, and J. C. Rasmussen, "Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate", *Journal of Lightwave Technology*, vol. 29, no. 17, pp. 2570–2576, 2011. DOI: 10.1109/JLT.2011.2160933.

[17] X. Liu, A. Chraplyvy, P. Winzer, R. Tkach, and S. Chandrasekhar, "Phase-conjugated twin waves for communication beyond the kerr nonlinearity limit", *Nature Photonics*, vol. 7, no. 7, pp. 560–568, 2013. DOI: 10.1038/NPHOTON.2013.109.