IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE
FOR PHYSICAL-LAYER WIRELESS COMMUNICATIONS

Received April 24, 2019, accepted May 20, 2019, date of publication May 30, 2019, date of current version June 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919983

A Model-Driven Deep Learning Method for LED Nonlinearity Mitigation in OFDM-Based Optical Communications

PU MIAO[1], (Member, IEEE), BINGCHENG ZHU[2], (Member, IEEE), CHENHAO QI[2], (Senior Member, IEEE), YI JIN[3], AND CHONG LIN[4], (Senior Member, IEEE)

[1]School of Electronic and Information Engineering, Qingdao University, Qingdao 266071, China
[2]School of Information Science and Engineering, Southeast University, Nanjing 210096, China
[3]China Academy of Space Technology of Xi'an, Xi'an 710100, China
[4]Institute of Complexity Science, Qingdao University, Qingdao 266071, China

Corresponding author: Chenhao Qi (qch@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61801257, Grant 61871119, and Grant 61801377, in part by the National Natural Science Foundation of Jiangsu under Grant BK20161428, in part by the Shandong Provincial Natural Science Foundation under Grant ZR2019BF001 and Grant SAST2016072, and in part by the China Postdoctoral Science Foundation under Grant 2019M652322.

ABSTRACT The nonlinearity of light emitting diodes (LED) has restricted the bit error rate (BER) performance of visible light communications (VLC). In this paper, we propose model-driven deep learning (DL) approach using an autoencoder (AE) network to mitigate the LED nonlinearity for orthogonal frequency division multiplexing (OFDM)-based VLC systems. Different from the conventional fully data-driven AE, the communication domain knowledge is well incorporated in the proposed scheme for the design of network architecture and training cost function. First, a deep neural network (DNN) combined with discrete Fourier transform spreading (DFT-S) is adopted at the transmitter to map the binary data into complex I-Q symbols for each OFDM subcarrier. Then, at the receiver, we divide the symbol demapping module into two subnets in terms of nonlinearity compensation and signal detection, where each subnet is comprised of a DNN. Finally, both the autocorrelation of the learned mapping symbols and the mean square error of demapping symbols are taken into account simultaneously by the cost function for network training. With this approach, the LED nonlinearity and the interference introduced by the multipath channel can be effectively mitigated. The simulation results show that the proposed scheme exhibits better BER performance than some existing methods and further accelerates the training speed, which demonstrates the prospective and validity of DL in the VLC system.

INDEX TERMS Deep learning, visible light communication, LED nonlinearity, orthogonal frequency division multiplexing.

I. INTRODUCTION

Visible light communication (VLC) has been regarded as one promising green technology for high speed indoor wireless accessing [1] since it can provide worldwide license-free bandwidth and high signal-to-noise ratio (SNR) [2]. With the architecture of low-cost intensity modulation and direct detection (IM/DD) in VLC, the data carrying intensity waveform is usually modulated onto the light emitting diodes (LED) at the transmitter and detected directly by photodetector (PD) at the receiver. Recently, optical orthogonal frequency division multiplexing (O-OFDM) has been widely employed in VLC system due to its high spectral efficiency and resistance to inter symbol interference (ISI) resulting from the optical diffuse channel [3]. However, the OFDM is characterized by high peak-to-average power ratio (PAPR) thus its performance is more susceptible to the LED nonlinearity, i.e., the characteristics between output optical power and input current [4]. As the LED is driven by the OFDM signal with high PAPR, a large number of nonlinear distortions will be consequently produced [5]. Therefore, the LED nonlinearity should be mitigated so as to improve the transmission quality.

PAPR reduction is one of valid solutions for LED nonlinearity mitigation. Several PAPR reduction schemes have been

The associate editor coordinating the review of this manuscript and approving it for publication was Ning Zhang.

proposed in [6]–[13]. Amplitude clipping and filtering [7] is the simplest approach and is favorable to the PAPR reduction in OFDM-based VLC systems since it distorts the peaks of transmitting signal to a desired value deliberately. However, clipping noise is produced and the bit error rate (BER) performance will be further deteriorated. Selective mapping (SLM) and partial transmit sequence (PTS) are signal scrambling techniques [8]–[10] which can statistically improve the PAPR distributions without any signal distortions. The key idea of signal scrambling is that several independent candidate vectors are first generated and the one with the minimum PAPR is selected for transmission. Nevertheless, exhaustive search has to be performed to find out the optimum phase factors, and the side information must be known by the receiver to reconstruct the transmitted data. Tone reservation and injection are also traditional methods to deal with the PAPR problem [11]; however it is troublesome to choose the optimal subset with large number of subcarriers. Unitary matrix transformations, such as discrete Fourier transform spreading (DFT-S) [12] and Hadamard precoding [13], can lower the PAPR of OFDM signals without causing any distortions. Nevertheless, the unitary matrix transformation is sensitive to LED nonlinearity, thus a robust equalizer should be usually performed before the inverse transformation module to reduce the ISI.

Pre-distorter [14], [15] and post-distorter [16]–[19] can be deployed for nonlinearity compensation (NC) at transmitter and receiver, respectively. In [14], a digital predistortion (DPD) based on the Taylor expansion model is employed in VLC system. With a priori information of LED transfer function, a time domain DPD is devised in [15] to compensate the distorted OFDM symbols. However, the memory nonlinearity effect of LED is not taken into account. In fact, the carrier-density response of LED is dependent on the frequency components of the driving current. As the bandwidth of modulated signal increased, more profound memory effect is exhibited in LED [16]. The well-known memory polynomial (MP) and generalized MP are traditional but still popular models for DPD [17]. Nevertheless, only the diagonal kernels are employed in MP which are insufficient to characterize the memory behavior of LED. Recent investigations experimentally demonstrate that the memory nonlinearity of LED can be well described by using the full Volterra series (VS) and its simplified variants [20]. Various VS pre-distortion and post-distortion have been proposed in VLC system. However, due to the inherent structure of VS, large number of coefficients have to be used and the structure complexity is higher than the conventional linear compensator. In addition, the reproducing kernel Hilbert space [19] based on minimum symbol error rate criterion can be also used for devising the adaptive post-distorter. But it requires huge storage and has high polynomial computational complexity, just like the problem encountered by the VS schemes.

Deep learning (DL) [21], which is a kind of machine learning approach based on the neural network (NN), has recently regained tremendous since it has achieved the breaking through performance in computer vision, natural language processing and automatic speech recognition. What's more, DL has also been introduced into the wireless physical layer communications. A comprehensive introduction and overview of DL for physical layer communications can be found in [22]–[24]. For instance, the convolution NN [24] is proposed for radio modulation recognition and it achieves competitive accuracy over the conventional methods which are relied on expert features. Channel encoding and decoding using different DL architectures were investigated in [25]–[27], such as high-density parity check codes with fully connected (FC) deep neural network (DNN) decoder [25], linear block codes with recurrent NN decoder [26] and polar codes with partitioned NN [27]. In [28], a five-layer FC-DNN with an end-to-end learning manner is deployed in OFDM receiver for channel estimation and symbol detection. The results demonstrate that the FC-DNN has the ability to learn the characteristics of wireless channels and then corrects both the channel distortion and interference automatically. In [29], a model-driven DL receiver, which constructs the network topology based on expert knowledge, is proposed in OFDM system to recover the transmitting data. Simulation results indicate that the model-driven DL receiver [29] offers more accurate channel estimation and higher data recovery accuracy when comparing with the existing methods and FC-DNN. Besides, it also exhibits relatively faster convergence speed, which is mainly benefited from the expert knowledge involved in DL scheme. In [30], the whole communication link can be also represented by the concept of autoencoder (AE), and simulation results imply that the joint optimization of transmitter and receiver over a physical channel could be achieved by an end-to-end learning. Motivated by the mind in [30], a novel PAPR reduction scheme using an AE is proposed in OFDM system [31] to jointly minimize the PAPR and BER performance. However, the application scenarios in [30] and [31] focus on radio frequency (RF) systems, which may not achieve better performance when extended to OFDM-based VLC systems. In addition, an AE-based transceiver for multi-colored VLC system is developed in [32] to deal with the color crosstalk.

Inspired by the approaches in [28]–[32], in this paper, we formulated the LED nonlinearity mitigation problems as a DL task, and proposed an AE-based nonlinearity mitigation scheme, abbreviated as NC-net, for the OFDM-based VLC systems. The key idea is that the constellation mapping and signal detection module in OFDM are represented by a model-driven AE network. With an efficient learning and end-to-end performance optimization, the proposed NC-net can capture the system imperfections and then mitigate the nonlinearity from the received symbols automatically. In addition, the high PAPR of OFDM signals is also reduced since the training procedure jointly optimizes the overall NC-net. The main contributions of this paper are summarized as follows:

- To augment the signal processing ability, the expert knowledge of conventional communication systems, in terms of rigorous mathematical models, is well incorporated in designing the AE architecture and the loss function, which is the main difference compared with the scheme in [31], [32]. Accordingly, it is favorable for reducing the network complexity and accelerating the training phase.
- We considered the memory effect of LED in an IM/DD channel, and applied the deep AE network to OFDM-based VLC system. However, in most literatures [29]–[32], the linear channel is considered and the nonlinear impairments, especially the memory effect of actual devices, are neglected for applications of DL technology.
- We investigated the impact of both regularization parameter and training SNR on network training from the aspect of PAPR reduction and BER performance improvement.
- After the network was trained with an elaborate loss function, the proposed scheme can address the LED memory nonlinearity and recover the transmitted symbols directly.

Furthermore, simulation results show that the proposed NC-net demonstrated better overall performances than some competing methods.

The remaining of the paper is organized as follows. In Section II, we give the preliminaries in terms of the PAPR performance of OFDM, the nonlinearity of IM/DD channel and the AE application for communication system. In Section III, the detailed structure of the proposed NC-net is presented and then the well-designed loss function is used to train the network. Section IV presents the results and discussions. Section V concludes the paper finally.

Notations: Matrices are denoted by upper boldface letters (e.g., $\mathbf{F}$). $x(n)$ denotes the $(n+1)^{th}$ element of the column vector $\mathbf{x}$. Also, $a_l^{\mathcal{S}}$ accounts for the parameter of a in *l-th* layer of $\mathcal{S}$ network. In addition, $\Re\{\cdot\}$, $\mathbb{E}\{\cdot\}$, $(\cdot)^T$, $(\cdot)^*$, $|\cdot|$ and $\mathcal{Z}(\cdot)$ are employed to represent the real, the mathematical expectation, the transpose, the complex conjugate, the absolute and symmetric conjugate operators, respectively. Let $\|\cdot\|_p$ denotes the ℓ_p-norm and $\mathcal{N}\left(\mu, \sigma^2\right)$ is the Gaussian distribution with mean μ and variance σ^2. We use $\widehat{a}$ to represent the estimation of a.

II. PRELIMINARIES

A. VLC SYSTEM

The basic block diagram of OFDM-based VLC system is shown in Fig. 1. A direct current (DC) is applied to the bipolar basedband O-OFDM signal $x(n)$ to shift the negative part to positive region, and the biased electrical signal modulates the power intensity of the optical signal of LED. After an optical link transmission, the optical intensity is captured by a PD with direct detection. Then, the converted electrical signal is delivered to the OFDM demodulator for data recovery. It is worth noting that a multi-path optical link and ambient noise

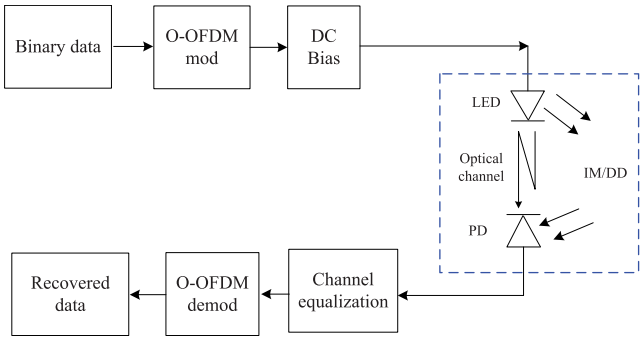

FIGURE 1. Block diagrams of O-OFDM-based VLC transceiver.

are also involved for an IM/DD channel. In VLC systems, the nonlinearity is mainly introduced by the LED, as driven by signals with high PAPR and high bandwidth. Besides, the PD is also a nonlinear device as the luminous intensity is large at the receiver. However, with the help of optical attenuator, the photocurrent of PD is directly proportional to the instantaneous optical power and its nonlinearity is not considered.

B. PAPR OF OFDM

Different with the RF-OFDM, the O-OFDM modulation generates only real-value signals after inverse fast Fourier transform (IFFT). In order to achieve real-valued time domain signals, the input frequency-domain symbol should fulfill Hermitian symmetry [7] and the IFFT size must be at least twice the subcarrier number. It is assumed that $\mathbf{X} = [X(0), X(1), \cdots, X(N-1)]^T$ is the symbol vector, then, the O-OFDM signal can be generated by

$$x(n) = \frac{1}{\sqrt{2N}} \sum_{k=0}^{2N-1} \Re\left\{C(k) \exp\left(\frac{j2\pi nk}{2N}\right)\right\}, \quad (1)$$

$$\mathbf{C} = \left[\mathbf{X}^T, (\mathcal{Z}(\mathbf{X}))^T\right]^T, \quad (2)$$

where N is the total subcarrier number, and $n = 0, \cdots, 2N-1$. The original $\mathbf{X}$ follows the Hermitian symmetry in (2). In order to eliminate the ISI, a certain length of cyclic prefix (CP) is inserted before the original $\mathbf{x}$. Since $\mathbf{X}$ to be statistically independent and identically distributed, the $\mathbf{x}$ follows Gaussian distribution $\mathcal{N}\left(0, \sigma_x^2\right)$ for sufficiently large N according to the central limit theorem.

Consequently, the PAPR of $\mathbf{x}$ can be defined as

$$\Upsilon_d = \frac{\max\limits_{0\le n\le 2N-1}\left\{|x(n)|^2\right\}}{\mathbb{E}\left\{|x(n)|^2\right\}} = \frac{\max\limits_{0\le n\le 2N-1}\left\{|x(n)|^2\right\}}{\sigma_x^2}. \quad (3)$$

Let $P_r(\cdot)$ denotes the probability and Υ_0 is the threshold, we can obtain

$$P_r\left(\frac{|x(n)|^2}{\sigma_x^2} \le \Upsilon_0\right) \approx \mathrm{erf}\left(\sqrt{\frac{\Upsilon_0}{2}}\right), \quad (4)$$

where $\mathrm{erf}(\psi) = 2/\sqrt{\pi}\int_0^{\psi} \exp\left(-t^2\right) dt$. The complementary cumulative distribution function (CCDF), which denotes

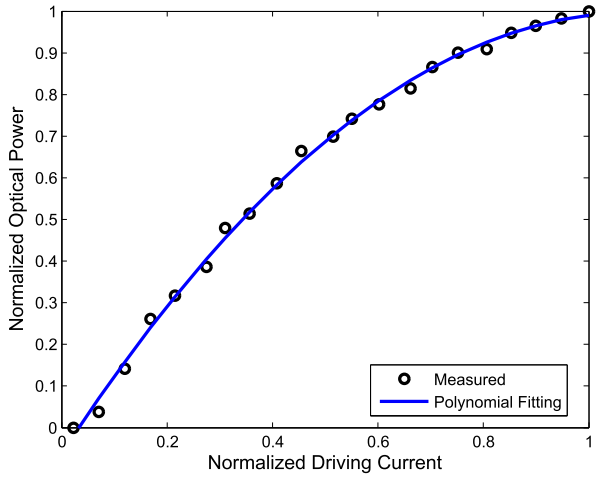

FIGURE 2. Static transfer function of a LED.

the probability of the PAPR that exceeds a given threshold during one OFDM symbol, is usually used to measure the distribution of the peak signals. Thus, the CCDF can be calculated by

$$\mathrm{CCDF} = P_r\left(\Upsilon_d > \Upsilon_0\right) = 1 - \mathrm{erf}^{2N}\left(\sqrt{\frac{\Upsilon_0}{2}}\right). \tag{5}$$

C. NONLINEAR IM/DD CHANNEL

An IM/DD channel of VLC systems should take into account the realistic effect of LED, optical wireless link and PD. In general, the LED nonlinearity primarily stems from the electrical-to-optical (E/O) conversation. Fig. 2 demonstrates the normalized static transfer function of a commercially available LED (Cree PLCC4) with different driving current, where the inputs and outputs are normalized by their maximum values, respectively. It can be clearly seen from the figure that the LED shapes the input driving current with strong nonlinear behavior. Meanwhile, the significant memory effect of LED is also arisen as the bandwidth of the injected current increased [20]. Although these frequency dependent nonlinearity could be represented by Volterra series in frequency domain or time domain, the Volterra structure is very complex and the number of coefficients increases exponentially with the raise of nonlinearity degree and memory length, leading to its impractical application for system simulations.

For simplicity, the Wiener model, which is a cascade of linear time-invariant (LTI) block and memoryless nonlinearity (NL) block, is used in many literatures [18] to characterize the LED behavior. For a Wiener model, the first LTI block employed a low-pass finite impulse response (FIR) filter to describe the memory effect in LED, and the second NL block represented the static transfer function of E/O conversation. In this paper, the impulse response of the LTI block is modeled by

$$h_{LTI}\left(n\right) = \exp\left(-2\pi n f_{3dB}\right), \tag{6}$$

where f_{3dB} is the normalized 3-dB cutoff frequency. The transfer function of NL block is modeled based on the measurements in Fig. 2, expressed as

$$h_{NI}\left(x\right) = -0.8940x^2 + 1.9472x - 0.0628. \tag{7}$$

In addition to the nonlinearity of LED, the multipath propagation environment should be also considered in the IM/DD channel. Here, we employed a ray-tracing-based approach [33] to get close to the realistic multipath propagation in VLC channel. After the calculations of the detected power and path lengths from source to detector for each ray, the channel impulse response can be expressed by

$$h_{\mathrm{VLC}}\left(n\right) = \sum_{i=1}^{N_r} P_i \delta\left(n - \tau_i\right), \tag{8}$$

where N_r is the number of received rays at the PD, $\delta\left(n\right)$ is the Dirac function, P_i and τ_i denote the optical power and propagation time of the *i-th* ray, respectively. Here, we consider the PD as a Dirac channel thus the corresponding impulse response can be presented by

$$h_{PD}\left(n\right) = R_{PD}\delta\left(n\right), \tag{9}$$

where R_{PD} is the responsitivity. The dominant ambient noise ε in an IM/DD channel includes shot noise and thermal noise of the PD, which can be modeled as additive white Gaussian noise (AWGN) with the variance of σ_ε^2. Therefore, the effective outputs of IM/DD channel can be expressed by follows

$$\mathbf{y} = \mathbf{h}_{NI}\left\{\mathbf{x} \otimes \mathbf{h}_{LTI}\right\} \otimes \mathbf{h}_{\mathrm{VLC}} \otimes \mathbf{h}_{PD} + \varepsilon, \tag{10}$$

where $\otimes$ denotes the convolution operation. In general, the nonlinear distortions in IM/DD channel can be solved independently by deploying the PAPR reduction block at transmitter or the NC block at receiver, but the global optimal performance of the entire transmission link still cannot be guaranteed by suboptimization for each module. Therefore, it hold promise for the overall performance improvement in VLC system if the suboptimization of the aforementioned two blocks were replaced by optimizing the end-to-end performance.

D. AE FOR COMMUNICATION SYSTEM

A typical structure of an AE is illustrated in Fig. 3, which contains an input layer, a hidden layer and an output layer. In addition, multiple dense layers or other network architecture, such as recurrent NN, can be also deployed as long as the encoder and decoder could handle the whole data sequence with deeper learning compatibility. As shown in this figure, AE is a special type of DNN that is used to learn a compressed representation form of input data $\mathbf{t}$, and trained with an unsupervised DL algorithm to produce the same values $\hat{\mathbf{t}}$ as $\mathbf{t}$ at the output layer. Therefore, the target output of the AE is the input itself. As we known, the main goal of the communication is to reconstruct the transmitted messages at the receiver sides, which can be recast as an end-to-end

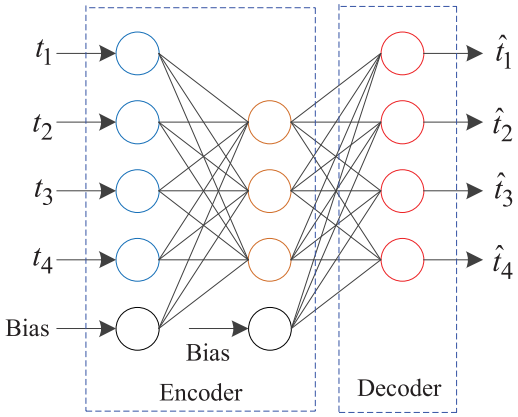

FIGURE 3. A simple AE consisting of an input layer, a hidden layer and an output layer.

reconstruction optimization task. Since the AE is trained to copy its input to its output, therefore, the AE could represent the entire communication link and optimize the transmitter and receiver jointly over the actual physical channel.

From a DL point of view, the transmitter and receiver of OFDM system can be interpreted as a particular of AE. Examples of AE-based OFDM system were discussed in [30]–[32], where the modules of I-Q symbols mapping and demapping in OFDM system are replaced by a single AE. In those schemes, both the encoder and decoder are consisted of a straightforward DNN. Through a DL method, the constellation mapping and demapping on each subcarrier are learned with the objective training function. However, the hardware impairments and expert knowledge of practical communication system are not well considered, which indicates that the aforementioned AE structure cannot be directly used in VLC system, especially involving the LED with memory nonlinear scenario. If we explored appropriate AE schemes to learn the effective data codings from the nonlinear contaminated signals, the LED nonlinearity would be well mitigated and the transmitted messages could be recovered correctly.

III. PROPOSED SCHEME

In this section, a model-driven AE, abbreviated as NC-net, is applied to the OFDM-based VLC system to optimize the end-to-end performance. The structure of the proposed NC-net, which embeds the encoder and decoder to a conventional OFDM system and incorporates the communication expert knowledge, is firstly analyzed. Then, the aperiodic autocorrelation of the learned I-Q samples is explored and introduced into the loss function for network training. After that, we train the proposed NC-net with piecewise constant learning rate for accelerating the training phase. In what follows, we assumed synchronization is perfectly obtained at the receiver.

A. SYSTEM ARCHITECTURE

A brief illustration of the proposed NC-net for VLC system is depicted in Fig. 4. Compared with the conventional OFDM, the encoder which contains one subnet $\mathcal{S}1$ and one DFT-spread module is employed at the transmitter to map the input bits into the I-Q constellations, and the decoder utilizes two subnets (i.e., $\mathcal{S}2$ and $\mathcal{S}3$) at the receiver for transmission impairments compensation and symbol detection, respectively. In addition, both the subnets $\mathcal{S}1$ and $\mathcal{S}3$ are composed of the cascaded sub-layer which contains dense

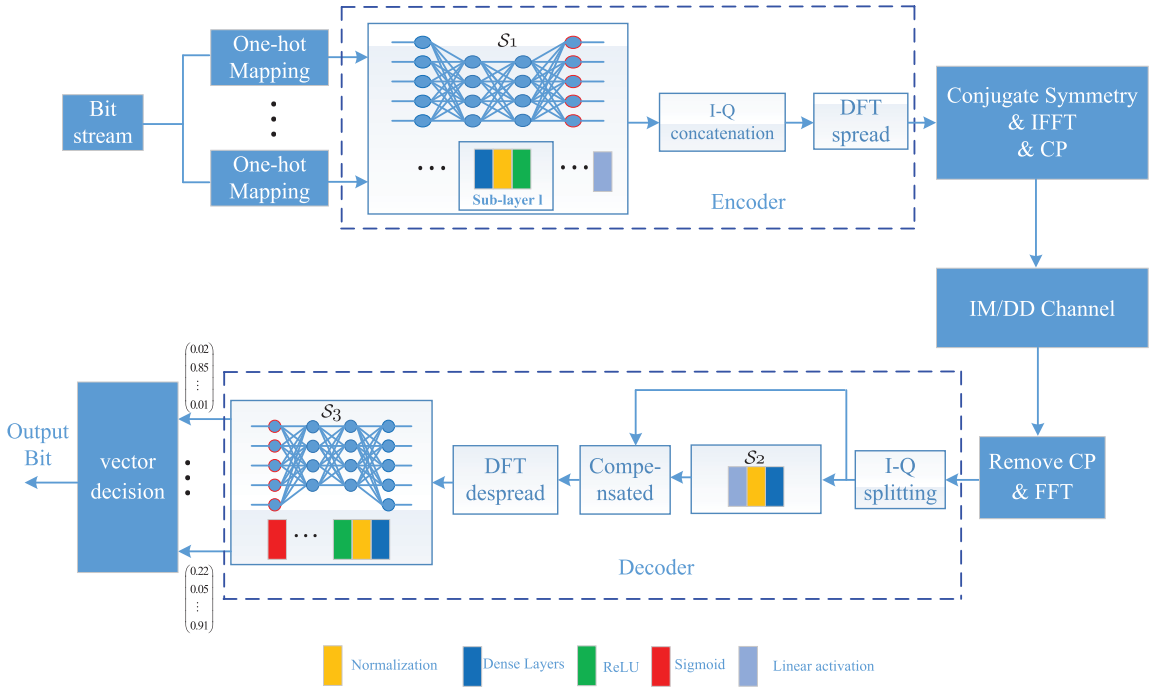

FIGURE 4. Diagram of the proposed model-driven NC-net for OFDM-based VLC system.

layer, normalization and activation function. Furthermore, the dropout layer can be also deployed after the dense layer for addressing over-fitting problem and improving generalization ability of the DNN model, especially when the number of sub-layers and hyper-parameters are very large. Therefore, the proposed NC-net are employed to replace parts of classic models of conventional OFDM system and to achieve the optimization of the end-to-end performance. It should be mentioned that the $\mathcal{S}1$, $\mathcal{S}2$ and $\mathcal{S}3$ are trained simultaneously with IM/DD channel and different amounts of neurons can also be employed in different sub-layers for reducing the network complexity.

At the transmitter, the serial bit stream are firstly converted to parallel symbol and then transformed into one-hot vector $\mathbf{1}_s \in \mathcal{U}^M$ (i.e., the *s-th* element of the M dimensional vector $\mathbf{1}_s$ is equal to one and other elements are zeros), where $\mathcal{U}^M$ is the one-hot mapping set and M is the modulation level. E.g., for $M = 4$, the symbol *'00'* could be represented as *'0001'*, and symbol *'01'* is *'0010'*, *etc*. Then, multiple $\mathbf{1}_s$ are recombined as data information vector $\mathbf{t} \in \mathbb{R}^{\mathcal{P}}$ ($\mathcal{P} = MN$) and fed into the encoder. Let $\mathcal{D}_l^{\mathcal{S}1}$ denotes the number of neurons, and $\mathbf{t}_l$ be the input for the l-th dense layer of $\mathcal{S}1$, then the outputs can be expressed as

$$\mathbf{h}_l^{\mathcal{S}1} = \mathbf{W}_l^{\mathcal{S}1}\mathbf{t}_l + \mathbf{b}_l^{\mathcal{S}1}, \tag{11}$$

where $\mathbf{W}_l^{\mathcal{S}1} \in \mathbb{R}^{\mathcal{D}_l^{\mathcal{S}1}\times\mathcal{P}}$ and $\mathbf{b}_l^{\mathcal{S}1} \in \mathbb{R}^{\mathcal{D}_l^{\mathcal{S}1}}$ are the weight matrix and bias vector for the l-th dense layer, respectively. Then, $\mathbf{h}_l^{\mathcal{S}1}$ passes through the normalization unit to normalize the input of activation function and to keep the same distribution for the input of each sub-layer during the NC-net training. In this paper, the Batch normalization is adopted and the corresponding outputs can be calculated as

$$\left(\mathbf{h}_l^{\mathcal{S}1}\right)_{Bat} = \alpha_l^{\mathcal{S}1}\frac{\mathbf{h}_l^{\mathcal{S}1} - \mathbb{E}\left\{\mathbf{h}_l^{\mathcal{S}1}\right\}}{\sqrt{\sigma^2_{\mathbf{h}_l^{\mathcal{S}1}} + \zeta}} + \beta_l^{\mathcal{S}1}, \tag{12}$$

where $(\cdot)_{Bat}$ denotes the output of Batch normalization, $\alpha_l^{\mathcal{S}1}$ and $\beta_l^{\mathcal{S}1}$ are the scaling and shift factors, respectively. In addition, ζ is always set to a number close to zero for preventing denominator from being zero. It is worth noting that the appropriate $\alpha_l^{\mathcal{S}1}$ and $\beta_l^{\mathcal{S}1}$ can be also learned in the training phase. Then, the $\left(\mathbf{h}_l^{\mathcal{S}1}\right)_{Bat}$ is fed into the activation function $\rho_l(\cdot)$ to produce the outputs of l-th sub-layer. We employ the ReLU activation to produce the same values as the $\left(\mathbf{h}_l^{\mathcal{S}1}\right)_{Bat}$ if only $\left(\mathbf{h}_l^{\mathcal{S}1}\right)_{Bat}$ is a positive real value. But beyond that, the linear activation is used for the last sub-layer due to the main goal of the final outputs of $\mathcal{S}1$ net is to produce the I-Q samples from the learned constellation plane.

Assuming $\mathcal{L}1$ layers was used for $\mathcal{S}1$, the output of $\mathcal{S}1$ can be expressed by

$$\mathbf{T} = \rho_{\mathcal{L}1}\left\{\left(\mathbf{h}_{\mathcal{L}1}^{\mathcal{S}1}\right)_{Bat}\right\}. \tag{13}$$

We divide the $\mathbf{T} \in \mathbb{R}^{2N}$ into two parts according to their odd and even index, and then concatenate them into complex I-Q symbol $\mathbf{X}$. In conventional OFDM scheme, the $\mathbf{X}$ is produced by the specific modulation methods, e.g., pulse amplitude modulation (PAM) or quadrature amplitude modulation (QAM). However, in our proposed scheme, those I-Q symbols on each subcarrier are generated by $\mathcal{S}1$ net through DL approach accompanied by minimizing the training cost, which will be discussed in the following sub-sections. Since the mapping patterns are learned from raw data automatically and the parameters of $\mathcal{S}1$ net are tuned to optimize end-to-end performance, the learned symbols may no longer have regular constellation patterns as the one of QAM or PAM. After that, $\mathbf{X}$ are spread by an $N \times N$ DFT matrix $\mathbf{F}$, and a new modulated symbol is generated by

$$X_{new}(k) = \sum_{n=0}^{N-1} X(n) F(n,k), \quad k = 0, \cdots, N-1, \tag{14}$$

where $F(n,k)$ accounts for the (n,k) element of matrix $\mathbf{F}$. Finally, the time domain O-OFDM signal $x_{new}(n)$ is obtained by using (1) and (2). After biased with DC, the $x_{new}(n)$ is fed into the IM/DD channel.

At the receiver, the optical signal is transformed into an electrical signal $\mathbf{y}$ by using a PD. For simplicity, R_{PD} is set to 1. After CP removal, FFT computation and symmetric data removal, the real and imaginary parts of received symbol $\mathbf{Y}$ are split and reformatted as a tensor $\bar{\mathbf{Y}}$. Instead of using a straightforward DNN to recover the transmitted information, we firstly applied the $\mathcal{S}2$ net for rough estimating the transmission impairments $\hat{\mathbf{H}}$, expressed as

$$\hat{\mathbf{H}} = \left(\mathbf{W}^{\mathcal{S}2}\bar{\mathbf{Y}} + \mathbf{b}^{\mathcal{S}2}\right)_{Bat}, \tag{15}$$

where $\mathbf{W}^{\mathcal{S}2}$ and $\mathbf{b}^{\mathcal{S}2}$ are the weight matrix and bias vector for $\mathcal{S}2$. It is noted that the input $\bar{\mathbf{Y}}$ is a $2N$-dimensional real-valued tensor. Additionally, $\mathcal{S}2$ may employ $\mathcal{L}2$ dense layers, and the Batch normalization and linear activation function should follow by the last sub-layer. In the compensation module, the traditional communication solution is employed and the compensated symbol can be roughly obtained by

$$\bar{\mathbf{X}} = \hat{\mathbf{H}}^{-1}\bar{\mathbf{Y}}. \tag{16}$$

Then, the $\bar{\mathbf{X}}$ is operated by the inverse DFT matrix in order to obtain the de-spread symbol $\bar{\mathbf{X}}_{de}$. After that, $\bar{\mathbf{X}}_{de}$ will be fed into $\mathcal{S}3$ net for refining the solution results of (16) while detecting the original symbols. The $\mathcal{S}3$ involves $\mathcal{L}3$ layers and its input is the concatenation of the real and imaginary parts of $\bar{\mathbf{X}}_{de}$. In addition, $\mathcal{S}3$ has the similar network architecture as $\mathcal{S}1$, but the main difference is that the activation function of last sub-layer is Sigmoid due to the main objective of $\mathcal{S}3$ is to recover the original binary data. Let φ_l, $\mathbf{W}_l^{\mathcal{S}3}$ and $\mathbf{b}_l^{\mathcal{S}3}$ denote the activation function, the weights and bias for the l-th dense layer of $\mathcal{S}3$, respectively. Then, the final output of $\mathcal{S}3$ can be expressed by

$$\hat{\mathbf{t}} = \varphi_{\mathcal{L}3}\left\{\left(\mathbf{W}_{\mathcal{L}3}^{\mathcal{S}3}\mathbf{t}_{\mathcal{L}3} + \mathbf{b}_{\mathcal{L}3}^{\mathcal{S}3}\right)_{Bat}\right\}, \tag{17}$$

where $\hat{\mathbf{t}} \in \mathbb{R}^{\mathcal{P}}$ is the approximate hot vector. Then, we can extract per M bits $\hat{t}_M$ from $\hat{\mathbf{t}}$ for symbol decision,

illustrated as

$$\hat{\mathbf{t}}_0 = \arg \min_{\substack{\mathcal{U}_i \\ i=1,2,\cdots,M}} \left| \hat{t}_M - \mathcal{U}_i \right|. \tag{18}$$

Finally, the symbol $\hat{\mathbf{t}}_0$ is transformed into the original binary bit stream. Therefore, $\mathcal{S}3$ not only refines the rough output results of $\mathcal{S}2$, but also obtains more accurate estimation of the transmitted data.

As aforementioned, the nonlinear noise and channel interference could distort the OFDM signal over the IM/DD transmission. To eliminate the noise and recover the original information, the proposed NC-net aims to learn features from raw data automatically and to find out an effective symbol mapping and demapping strategy.

B. NETWORK TRAINING

During the training stage, the proposed NC-net is trained to optimize the end-to-end performance via tuning the parameters $\Theta = \left\{\mathbf{W}^{\mathcal{S}1}, \mathbf{b}^{\mathcal{S}1}, \mathbf{W}^{\mathcal{S}2}, \mathbf{b}^{\mathcal{S}2}, \mathbf{W}^{\mathcal{S}3}, \mathbf{b}^{\mathcal{S}3}\right\}$ so that the reconstruction $\hat{\mathbf{t}}$ is closer to the raw $\mathbf{t}$. The relation between $\mathbf{t}$ and $\hat{\mathbf{t}}$ can be measured with the mean square error (MSE), which can be demonstrated by

$$\mathrm{MSE}\left(\mathbf{t}, \hat{\mathbf{t}}\right) = \left\| \mathbf{t} - \hat{\mathbf{t}} \right\|_2^2. \tag{19}$$

The MSE is usually used for loss function in the network training. However, as previously mentioned, the high PAPR will introduce more LED nonlinearity thus degrade the MSE. This inspires us that if we could introduce the PAPR-related items into the loss function, the network will be also trained to reduce the PAPR, which is beneficial to the achievement of the objective MSE. In other words, the speed of network training would be accelerated to a certain extent with the help of the PAPR-related items. Therefore, the BER and PAPR performance should be taken into the account simultaneously during the network training of NC-net.

On the other hand, the (3) with an unit mean power ($\sigma_x^2 = 1$) can be modified as

$$\begin{aligned}
\Upsilon_d &= \frac{\max\limits_{0 \leqslant k \leqslant 2N-1} \left\{ |x(n)|^2 \right\}}{\sigma_x^2} \\
&= \frac{1}{2N} \sum_{p=0}^{2N-1} \sum_{q=0}^{2N-1} C(p)\, C^*(q) \exp\left(\frac{j2\pi (p-q) n}{2N} \right) \\
&= \frac{N+2}{2N} \Re \left\{ \sum_{q=1}^{2N-1} \exp\left(\frac{j2\pi qn}{2N} \right) \sum_{p=0}^{2N-1-q} C(p+q)\, C^*(p) \right\} \\
&= \frac{N+2}{2N} \Re \left\{ \sum_{q=1}^{2N-1} \rho_C(q) \exp\left(j\frac{2\pi qn}{2N} \right) \right\},
\end{aligned} \tag{20}$$

where $\rho_C(q) = \sum\limits_{p=0}^{2N-1-q} C(p+q)\, C^*(p)$ is the aperiodic autocorrelation function (ACF) of $\mathbf{C}$. For the complex symbol $\mathcal{X}$, we can obtain the followings inequalities

$$\Re\{\mathcal{X}\} \leqslant |\mathcal{X}|, \quad \left| \sum_{n=0}^{N-1} \mathcal{X} \right| \leqslant \sum_{n=0}^{N-1} |\mathcal{X}|. \tag{21}$$

Therefore, the Υ_d satisfies

$$\Upsilon_d \leqslant \frac{1}{2} + \frac{1}{N} \sum_{q=1}^{2N-1} |\rho_C(q)| = \frac{1}{2} + \frac{1}{N} \xi_{\mathbf{C}}, \tag{22}$$

where $\xi_{\mathbf{C}} = \sum\limits_{q=1}^{2N-1} |\rho_C(q)|$ is the sum of ACF. According to (22), we know that a close relationship indeed exists between Υ_d and $\xi_{\mathbf{C}}$. For a fixed N, the maximum value of Υ_d is limited by the value of $\xi_{\mathbf{C}}$, therefore, the more reduction of $\xi_{\mathbf{C}}$ means a lower Υ_d can be indirectly obtained. As all the elements of vector $\mathbf{C}$ are the same, the Υ_d can achieve its maximum value. As a result, the $\xi_{\mathbf{C}}$ will be chosen as the PAPR-related items and added into the loss function for network training.

Considering (19) and (22), we can design the training cost of NC-net as followings

$$\mathcal{J}\left(\mathbf{t}, \hat{\mathbf{t}}\right) = \mathrm{MSE}\left(\mathbf{t}, \hat{\mathbf{t}}\right) + \eta \xi_{\mathbf{C}}(\mathbf{t}). \tag{23}$$

Here, η is a regularization parameter that controls the equilibrium relationship between BER and PAPR performance in the network training. The main goal is to learn the efficient $\hat{\Theta}$ that minimize the objective $\mathcal{J}\left(\mathbf{t}, \hat{\mathbf{t}}\right)$ for each training samples, expressed as

$$\hat{\Theta} = \arg \min_{\Theta} \mathcal{J}\left(\mathbf{t}, \hat{\mathbf{t}}\right). \tag{24}$$

Moreover, some advanced backpropagation methods with a favorable learning rate λ can be utilized. For comprehensive consideration of computational efficiency and stability, the adaptive moment estimation (Adam) optimizer with the piecewise learning rate are adopted in this paper. It is noted that the OFDM symbols should be randomly generated during each training epoch so that we can collect the diverse and abundant training set, which is favorable for parameters learning. Meanwhile, the training SNR Λ_{Tr} is fixed to a constant value in a single training phase.

C. COMPLEXITY ANALYSIS

The encoder and decoder are dominant in VLC transceiver with regarding to the complexity of the proposed NC-net. However, the computational complexity of the training phase seems difficult to quantify. In most case, the DL network is usually trained by an off-line way. As long as the Θ is determined, only some adders and multipliers are needed in the forward propagation of NC-net. Therefore, the computational complexity of the forward propagation is considered here. Assume all the hidden dense layers have $\mathcal{D}$ neurons, the computation complexity of $\mathcal{S}1$ can be approximately expressed by $\mathcal{O}\left((M+2)N\mathcal{D} + (\mathcal{L}1-2)\mathcal{D}^2\right)$. Similarly, the complexity of $\mathcal{S}2$ and $\mathcal{S}3$ are expressed as $\mathcal{O}\left(4N\mathcal{D} + (\mathcal{L}2-2)\mathcal{D}^2\right)$

TABLE 1. Parameters of the proposed NC-net.

Parameters	Values
Dimensions for $\mathcal{S}1$	
Input	MN
Dense (ReLU)	256
Dense	$2N$
Dimensions for $\mathcal{S}2$	
Input	$2N$
Dense (ReLU)	128
Dense	$2N$
Dimensions for $\mathcal{S}3$	
Input	$2N$
Dense (ReLU)	256
Dense (Sigmoid)	MN
Training	
Optimizer	Adam
Learning rate	Piecewise
SNR	$5-40$ dB

and $\mathcal{O}\left((M+2)N\mathcal{D}+(\mathcal{L}3-2)\mathcal{D}^2\right)$, respectively. Moreover, two extra N-point FFT and IFFT modules with $\mathcal{O}\left(N\log_2(N)\right)$ are also used to spread and de-spread the OFDM symbols. As for the computational complexity, these above factors should be taken into account jointly.

IV. RESULTS AND DISCUSSIONS

In this section, simulations of the proposed NC-net over IM/DD channel are conducted and the system performances in terms of the PAPR and BER are investigated. For simplicity, the O-OFDM symbol with total 64 subcarriers are considered, where the modulation levels $M = 4$. In addition, an 256-point IFFT/FFT is also used to produce the time domain signal. Assuming the LED model in (6) with 3-dB cutoff frequency of 20 MHz and the normalized DC is fixed to 0.25. Furthermore, the indoor VLC channel recommended by IEEE 802.15 is adopted [33] and the multi-path delay spread is set to 12ns. In most cases, the rate of reflected light is small and the line-of-sight path are dominant in the impulse response. Therefore, the influence of LED nonlinearity is large and it depends on system performance greatly. It should be noted that the convolution behavior of LED and multipath propagation can be realized by the *tf.nn.conv1d* function in *TensorFlow* based on the appropriate dimensionality conversion. In order to combat with the ISI from the channel, the CP ratio is designed as 1/8. The parameters involving $\mathcal{S}1$, $\mathcal{S}2$ and $\mathcal{S}3$ are shown in Table 1.

In the training phase, 500 training samples are used for each epoch and total 300,000 epochs are implemented for a fixed Λ_{Tr}. The learning rate can be set as a piecewise function with initial value of 0.01 and least value of 1×10^{-6}. It decreased 5-fold every 20,000 epochs. In addition, the training procedure is implemented in *TensorFlow* and the test set containing 8,000 samples are employed in the testing procedure. If the total BER is lower than 1×10^{-3}, we think that the network training is successful because this BER could be corrected to value $\leqslant 10^{-9}$ with the help of forward error correction coding, which is not implemented here.

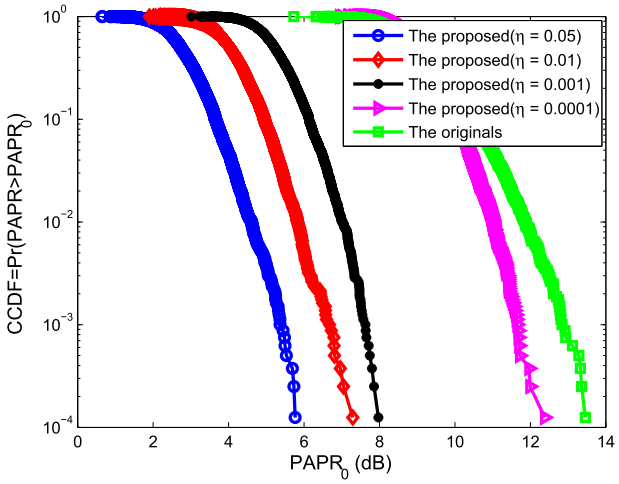

FIGURE 5. PAPR performance of the proposed scheme under different regularization parameter.

A. IMPACT OF REGULARIZATION PARAMETER

The appropriate regularization parameter η_{opt} should be firstly determined so that it can establish good base for improving the holistic capability of the proposed scheme. As we known, the LED owns the limited linear range thus the high PAPR of driving signal will introduce more signal distortions into the system. That is to say, the existence of LED brings about the restrictive relation between the PAPR and BER. Therefore, the LED model in (6) and (7) are not considered here so that we can investigate the real effect of the regularization parameter η on the PAPR and BER performance, independently. To verify the impact of η, the proposed NC-net is trained under different η. Additionally, the training SNR is set to $\Lambda_{Tr} = 20$ dB. After the network training, the CCDF and BER are then calculated.

Fig. 5 presents the CCDF curves of the proposed scheme as η varies from 0.0001 to 0.05. It can be observed that the proposed scheme exhibits a lower PAPR as compared with the conventional scheme, which indicates that the peaks of O-OFDM can be reduced efficiently through learning of the constellation mapping patterns. Besides, the bigger η will lead to better PAPR performance during the network training.

Fig. 6 shows the corresponding BER performance without considering the LED transforming. In addition, the idea case where the multipath channel fading has been perfectly compensated is also presented. As seen in the figure, the BER are gradually decayed as the regularization parameter η increasing, just contrary to the trend of CCDF. When the η is set to a small value, the NC-net is trained for improving the BER whereas putting less effort into reducing the PAPR. Furthermore, the training epochs will be decreased for small η. For instance, in order to achieve the given MSE $= 1 \times 10^{-3}$, approximately 103,600 epochs are consumed for the training with $\eta = 0.001$ whereas 86,300 epochs will be cost for $\eta = 0.0001$. However, as for the cases of $\eta = 0.05$ and $\eta = 0.01$, there is still a long range to approach the BER of 1×10^{-3}. Therefore, the appropriate η_{opt} can provide

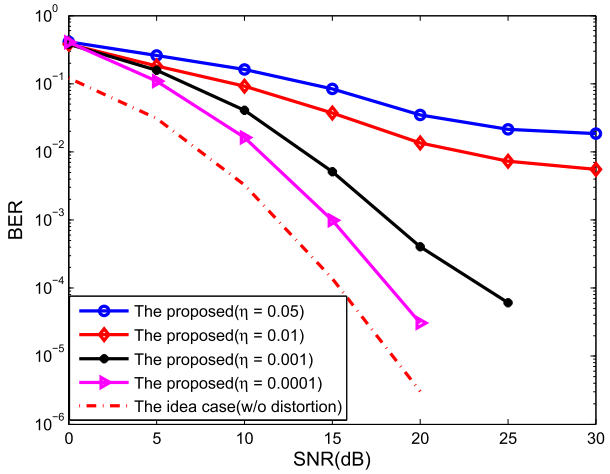

FIGURE 6. BER performance of the proposed scheme under different regularization parameter.

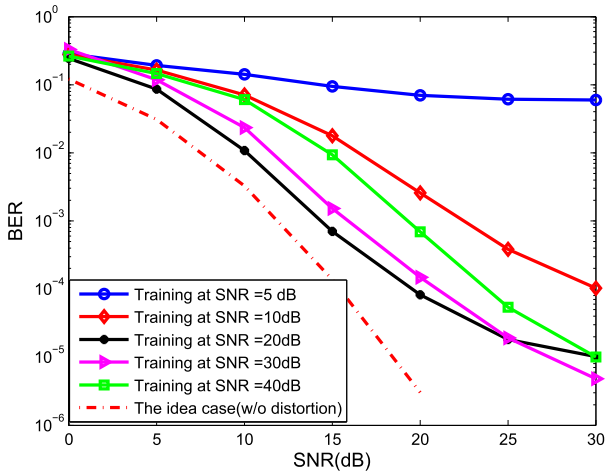

FIGURE 7. BER performance of the proposed scheme under different training SNR.

favorable trade-offs among the PAPR and BER performance and can accelerate the network training. It is worth noting that the appropriate η_{opt} should be carefully investigated according to the modulation parameters. In this paper, $\eta_{opt} = 0.001$ is employed in the following simulations.

B. IMPACT OF TRAINING SNR

The quality of the training set will affect the training efficiency and output accuracy of the network. For a communication task, the favorable SNR of the training samples should be carefully considered so that the proposed NC-net could be trained effectively under the proper noise level. As a result, the proposed NC-net is trained with Λ_{Tr} varying from 5 to 40 dB, respectively. The corresponding BER performance is depicted in Fig. 7. Mentioned that the LED model is involved in the network training.

As clearly shown in the figure, the proposed NC-net can hardly work normally in the case of $\Lambda_{Tr} = 5$ dB because the neural network cannot learn the useful features of modulated symbols in a big noisy environment. As the Λ_{Tr} increased, the BER performance is getting better and the according

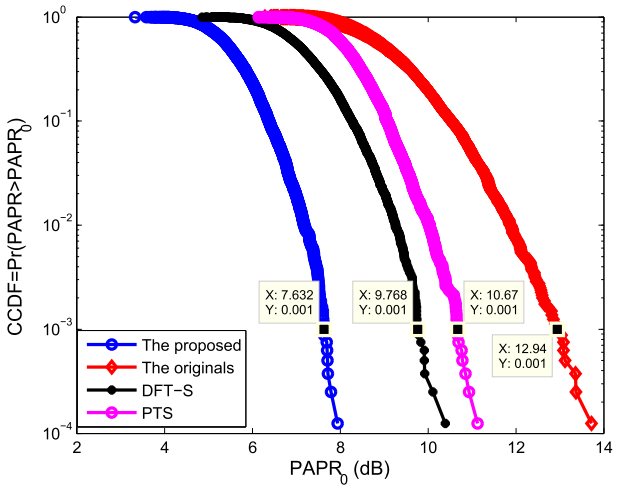

FIGURE 8. The CCDF performance comparison among different schemes.

curves are gradually approaching the idea's, which indicates that the LED nonlinearity had been compensated by the proposed NC-net to some extent. Even though the BERs of the NC-net are not exactly same with the idea case, the SNR gap is relatively smaller at a BER of 1×10^{-3}, especially for the network trained at $\Lambda_{Tr} = 20$ dB, e.g., the SNR gap is 2.1 dB as compared to the idea case. However, the BERs are degraded when the higher Λ_{Tr} are employed. As seen the curve for $\Lambda_{Tr} = 40$ dB, the BER performance is even inferior to that of $\Lambda_{Tr} = 30$ dB. In that case, the NC-net is confined to a very small region during the training phase, which leads to inadequate learning of the environmental noise and obscuring the signal characteristics around the objective constellation points. In fact, we should choose the favorable training SNR according to the modulation level so that the environmental noise would not confuse the symbol judgment region. Therefore, $\Lambda_{Tr} = 20$ dB is employed in our network training based on the above analysis.

C. PAPR PERFORMANCE

For comparison, the original O-OFDM and two well-known PAPR reduction schemes, i.e., PTS and DFT-S, are also investigated in this part. For PTS algorithm, the input symbol vector is interleaved into 4 disjoint sub-blocks and 2 phase factors are employed. Fig. 8 demonstrates the CCDF performances of these 4 schemes based on 9,000 random OFDM symbols. From the figure, we find that the proposed scheme has much lower PAPR and it respectively achieves 2.14, 3.04 and 5.31 dB performance gains than that of the other 3 schemes at CCDF $= 10^{-3}$. Even the CCDF of DFT-S exceeds that of PTS and original O-OFDM, but it is inferior to that of the proposed scheme because the NC-net had been integrated the DFT-S technique jointly into the deep learning network.

The probability density functions (PDFs) of the sum of ACF with respect to these 4 schemes are demonstrated in Fig. 9. Let ξ_{C_NC} denotes the sum of ACF for the proposed NC-net. As seen in the figure, all of the PDF curves

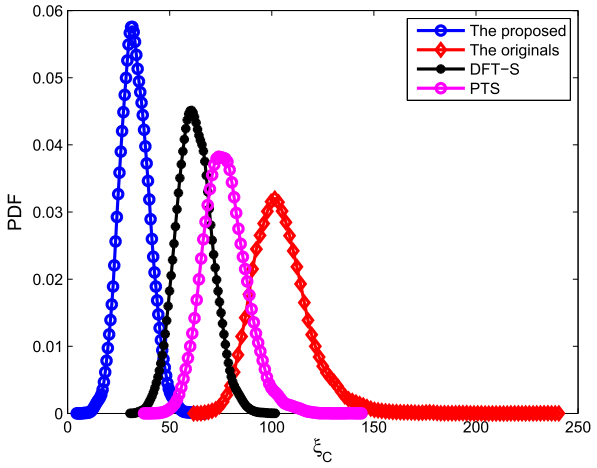

FIGURE 9. The PDF of the ACF among different schemes.

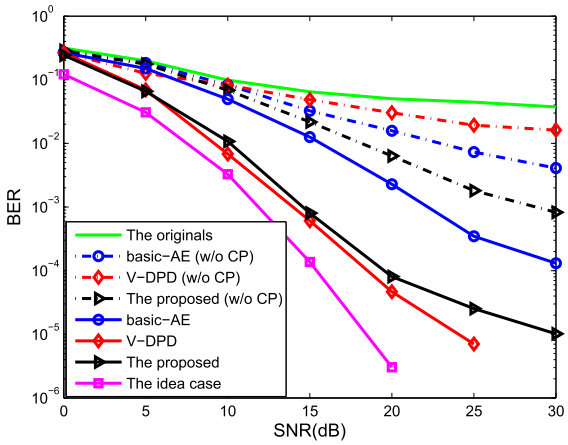

FIGURE 10. The BER performance comparison among different schemes.

have similar shape with approximately Gaussian distribution. Those curves of the PTS, DFT-S and proposed scheme move a great distance away from the originals. As we known, the lower value of $\mathbb{E}\{\xi_{\mathbf{C}}\}$ implies a lower value of $\xi_{\mathbf{C}}$ for Gaussian distribution samples. Therefore, compared with the other 3 curves, the $\mathbb{E}\{\xi_{\mathbf{C\_NC}}\}$ is the lowest one which indicates that the proposed NC-net has indeed reduced the PAPR to the most extent. In addition, the more of $\xi_{\mathbf{C\_NC}}$ is reduced, the better PAPR performance is obtained by the NC-net. Thus, it can be concluded that the proposed NC-net outperforms the conventional schemes in terms of PAPR reduction.

D. BER PERFORMANCE

In this study, the BER performances of the proposed scheme in the case of transmissions with CP and without CP are investigated, respectively. For comparisons, the competing methods, in terms of the Volterra-based DPD (abbreviated to V-DPD) and the basic data-driven AE (abbreviated to basic-AE), are also conducted for LED nonlinearity compensation. With regard to V-DPD scheme, the memory length is set to 4 and the nonlinearity order is fixed to 6. As for basic-AE, both the encoder and decoder are comprised by the FC-DNN without any expert knowledge and the network dimension can be adopted as suggested in literature [31]. Meanwhile, the original case (i.e., contaminated by channel distortion but detecting without any compensation measures) and the idea case are also implemented for convenient analysis and comparison. As a note, the parameters of OFDM in each scheme remain the same.

With the testing SNR varying from 0 to 30 dB, the BER performance of these schemes is demonstrated in Fig. 10, where the solid lines represent the schemes with CP and the dashed lines denote the cases without CP. As shown in the figure, both the proposed and competing schemes had indeed improved the BER performance and had compensated the nonlinear distortion to a certain extent, as compared to the results of the original case. For these schemes involving CP, the V-DPD achieves the best scores because it employs the accurate modeling and rigorous analytical approach of the communication theory. Whereas the unfavorable factor for V-DPD is that the identification accuracy of model coefficients is susceptible to environmental interference and large computational complexity has to be cost. As for the proposed NC-net, it also exhibits the excellent BER obviously and achieves the significant performance improvement than that of basic-AE, e.g., for SNR = 20 dB, the BER had been reduced at least an order of magnitude by the NC-net as compared to the basic-AE. In addition, the curve is very close to that of V-DPD. These facts show that the non-linear noise caused by LED and the ISI from multipath had been effectively compensated by NC-net. As for the training convergence, the NC-net merely takes about 150,000 epochs to reach the MSE of 1×10^{-3}, nevertheless nearly 225,000 epochs is cost for basic-AE. The main reason for the difference lies in the fact that the basic-AE is designed without relying on any mathematical model and the hyper-parameters are just tuned by conducting the network training, resulting in a huge time overheads. This superior convergent ability benefits from expert-domain knowledge incorporated in the NC-net which can reduce the complexity and accelerate the training speed.

As we known, the CP is added before the OFDM symbols to mitigate the ISI caused by the transmission channel. If the CP length is shorter than channel delay spread, the BER performance would be degraded consequently. Clearly seen from the figure, all of the BERs have a certain degree of decline for the cases of CP removal. Furthermore, the BER of V-DPD (w/o CP) has decayed a lot from the V-DPD since they are working intrinsically based on model solving and do not have the ability to learn the channel feature from the external environment. Additionally, the V-DPD (w/o CP) method becomes saturated when the SNR is larger than 25 dB, and the basic-AE (w/o CP) is still struggling to approach the objective BER of 1×10^{-3}. However, the NC-net (w/o CP) has worked very well in resolving ISI and has reached to the BER of 1×10^{-3} at SNR = 30 dB, which indicates that the proposed scheme can learn the channel features availably and rectify

the negative effect of CP removal with the help of model-driven DL methodology.

In summary, we can conclude that the proposed NC-net is beneficial to the nonlinearity mitigation and outperforms some conventional methods from the above simulations and discussions.

V. CONCLUSION

In this paper, a model-driven NC-net combining DL with expert-domain knowledge has been proposed to mitigate the LED nonlinearity for OFDM-based VLC system. The simulation results show that the nonlinear distortions including the memory and memoryless nonlinearity are effectively mitigated and the ISI introduced by the multipath of IM/DD channel is also rectified by the proposed NC-net. With the incorporated expert-domain knowledge, the NC-net exhibits better BER performance than basic-AE method and converge faster in the training phase. Furthermore, the NC-net can still work effectively in the case of CP removal, which benefits from the powerful learning capability of DL and then overcomes the deficiencies in practical communications. This paper is an infancy study of the DL applications on physical layer communications. A further degree of rigorous analysis and comprehensive experiments are left for the future work, such as theoretical analysis of the model-driven NC-net need to be performed to fine-tune the DL model for achieving better BER performance, and for overcoming both the randomness and instability of network training.

ACKNOWLEDGMENTS

The authors would like to express gratitude to the editors and the anonymous reviewers for their insightful suggestions and general assistance.

REFERENCES

[1] X. Wu, M. Safari, and H. Haas, "Access point selection for hybrid Li-Fi and Wi-Fi networks," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5375–5385, Dec. 2017.
[2] W. Xu, J. Wang, H. Shen, H. Zhang, and X. You, "Indoor positioning for multiphotodiode device using visible-light communications," *IEEE Photon. J.*, vol. 8, no. 1, Feb. 2016, Art. no. 7900511.
[3] L. Chen, B. Krongold, and J. Evans, "Performance analysis for optical OFDM transmission in short-range IM/DD systems," *J. Lightw. Technol.*, vol. 30, no. 7, pp. 974–983, Apr. 1, 2012.
[4] H. Lu, Y. Hong, L.-K. Chen, and J. Wang, "On the study of the relation between linear/nonlinear PAPR reduction and transmission performance for OFDM-based VLC systems," *Opt. Express*, vol. 26, no. 11, pp. 13891–13901, May 2018.
[5] X. Deng, S. Mardanikorani, Y. Wu, K. Arulandu, B. Chen, A. M. Khalid, and J.-P. M. G. Linnartz, "Mitigating LED nonlinearity to enhance visible light communications," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5593–5607, Nov. 2018.
[6] L. Deng, Y. Fan, and Q. Zhao, "A novel PAPR reduction scheme for VLC DCO-OFDM systems," *Opt. Commun.*, vol. 426, pp. 164–169, Nov. 2018.
[7] P. Miao, C. Qi, L. Fang, K. Song, and Q. Bu, "Deep clipping noise mitigation using ISTA with the specified observations for LED-based DCO-OFDM system," *IET Commun.*, vol. 12, no. 20, pp. 2582–2591, Dec. 2018.
[8] V. Sudha, M. Syamkumar, and D. S. Kumar, "A low complexity modified SLM and companding based PAPR reduction in localized OFDMA," *Wireless Pers. Commun.*, vol. 96, no. 2, pp. 3207–3226, Sep. 2017.
[9] H. Chen and K.-C. Chung, "A PTS technique with non-disjoint sub-block partitions in M-QAM OFDM systems," *IEEE Trans. Broadcast.*, vol. 64, no. 1, pp. 146–152, Mar. 2018.
[10] Y. A. Jawhar, L. Audah, M. A. Taher, K. N. Ramli, N. S. M. Shah, M. Musa, and M. S. Ahmed, "A review of partial transmit sequence for PAPR reduction in the OFDM systems," *IEEE Access*, vol. 7, no. 1, pp. 18021–18041, Feb. 2019.
[11] J. Hou, X. Zhao, F. Gong, F. Hui, and J. Ge, "PAPR and PICR reduction of OFDM signals with clipping noise-based tone injection scheme," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 222–232, Jan. 2017.
[12] Z. Wu, Y.-L. Gao, Z.-K. Wang, C. You, C. Yang, C. Luo, and J. Wang, "Optimized DFT-spread OFDM based visible light communications with multiple lighting sources," *Opt. Express*, vol. 25, no. 22, pp. 26468–26482, Nov. 2017.
[13] P. Miao, P. Chen, and Z. Chen, "Low-complexity PAPR reduction scheme combining multi-band Hadamard precoding and clipping in OFDM-based optical communications," *Electronics*, vol. 7, no. 2, p. 11, Jan. 2018.
[14] B. Inan, S. C. J. Lee, S. Randel, I. Neokosmidis, A. M. J. Koonen, and J. W. Walewski, "Impact of LED nonlinearity on discrete multitone modulation," *J. Opt. Commun. Netw.*, vol. 1, no. 5, pp. 439–451, Oct. 2009.
[15] H. Elgala, R. Mesleh, and H. Haas, "Non-linearity effects and predistortion in optical OFDM wireless transmission using LEDs," *Int. J. Ultra Wideband Commun. Syst.*, vol. 1, no. 2, pp. 143–150, Jan. 2009.
[16] G. Zhang, J. Zhang, X. Hong, and S. He, "Low-complexity frequency domain nonlinear compensation for OFDM based high-speed visible light communication systems with light emitting diodes," *Opt. Express*, vol. 25, no. 4, pp. 3780–3794, 2017.
[17] W. Zhao, Q. Guo, J. Tong, J. Xi, Y. Yu, P. Niu, and X. Sun, "Orthogonal polynomial-based nonlinearity modeling and mitigation for LED communications," *IEEE Photon. J.*, vol. 8, no. 4, Aug. 2016, Art. no. 7905312.
[18] H. Qian, S. Yao, S. Cai, and T. Zhou, "Adaptive postdistortion for nonlinear LEDs in visible light communications," *IEEE Photon. J.*, vol. 6, no. 4, pp. 1–8, Apr. 2014.
[19] R. Mitra and V. Bhatia, "Low complexity post-distorter for visible light communications," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 1977–1980, Sep. 2017.
[20] T. Kamalakis, J. W. Walewski, G. Ntogari, and G. Mileounis, "Empirical Volterra-series modeling of commercial light-emitting diodes," *J. Lightw. Technol.*, vol. 29, no. 14, pp. 2146–2155, Jul. 15, 2011.
[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[22] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2018.
[23] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Commun.*, vol. 14, no. 11, pp. 92–111, 2017.
[24] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.
[25] E. Nachmani, Y. Be'ery, and D. Burshtein, "Learning to decode linear codes using deep learning," in *Proc. 54th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, VA, USA, Sep. 2016, pp. 341–346.
[26] E. Nachmani, E. Marciano, D. Burshtein, and Y. Be'ery, "RNN decoding of linear block codes," 2017, *arXiv:1702.07560*. [Online]. Available: https://arxiv.org/abs/1702.07560
[27] S. Cammerer, T. Gruber, J. Hoydis, and S. ten Brink, "Scaling deep learning-based decoding of polar codes via partitioning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.
[28] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.
[29] X. Gao, S. Jin, C.-K. Wen, and G. Y. Li, "ComNet: Combination of deep learning and expert knowledge in OFDM receivers," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2627–2630, Dec. 2018.
[30] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. T. Brink, "OFDM-autoencoder for end-to-end learning of communications systems," in *Proc. 19th Int. Workshop Signal Process. Adv. Wireless Commun.*, Kalamata, Greece, Jun. 2018, pp. 1–5.
[31] M. Kim, W. Lee, and D.-H. Cho, "A novel PAPR reduction scheme for OFDM system based on deep learning," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 510–513, Mar. 2018.
[32] H. Lee, I. Lee, and S. H. Lee, "Deep learning based transceiver design for multi-colored VLC systems," *Opt. Express*, vol. 26, no. 5, pp. 6222–6238, 2018.
[33] M. Uysal, F. Miramirkhani, O. Narmanlioglu, T. Baykas, and E. Panayirci, "IEEE 802.15.7r1 reference channel models for visible light communications," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 212–217, Jan. 2017.