



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

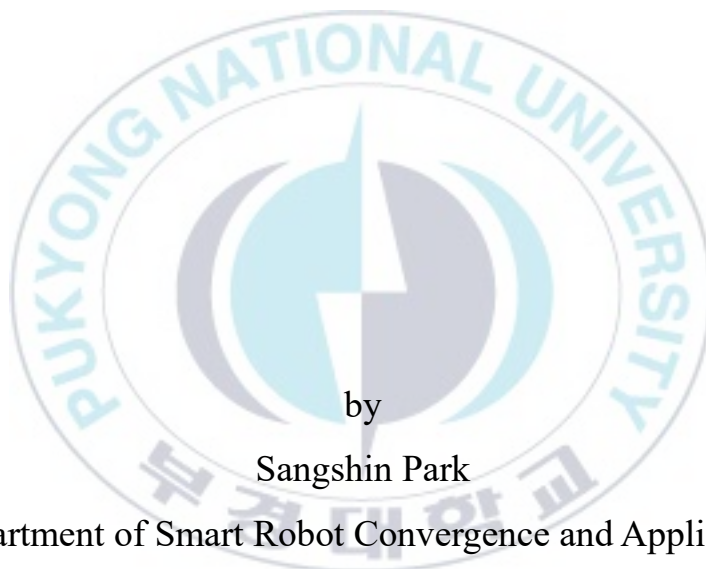
저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Thesis for the Degree of Master of Engineering

Deep Learning-Based Optical Wireless Communication System Design



by

Sangshin Park

Department of Smart Robot Convergence and Application

Engineering

The Graduate School

Pukyong National University

February, 2022

Deep Learning-Based Optical Wireless Communication System Design

딥 러닝 기반 광 무선 통신 시스템 설계

Advisor: Prof. Hoon Lee

by
Sangshin Park

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
in Department of Smart Robot Convergence and Application Engineering,
The Graduate School,
Pukyong National University
February, 2022

Deep Learning-Based Optical Wireless Communication System Design

A dissertation

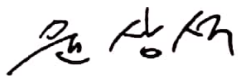
by

Sangshin Park

Approved by:



Professor Jun-Pyo Hong,
(Chairman)



Professor Sang-Seok Yun,
(Member)



Professor Hoon Lee,
(Member)

February 25, 2022

Contents

List of Figures.....	ii
List of Tables.....	ii
Abstract.....	iii
I. Introduction.....	1
II. Data Generation.....	6
2.1 Transmitter.....	7
2.2 Receiver.....	8
III. Proposed CNN-based OCC Receiver	10
IV. Simulation results.....	15
V. Conclusions	22
References.....	23

Figure List

Fig. 1.	Experimental environment.....	6
Fig. 2.	Transmit LED matrix and OOK modulation strategy.....	7
Fig. 3.	Sample images of training dataset.....	9
Fig. 4.	Proposed CNN structure.....	12
Fig. 5.	Convergence behavior for various learning rate values.....	16
Fig. 6.	Convergence behavior for various kernel configurations...	17
Fig. 7.	Convergence behavior for various layer setups.....	18
Fig. 8.	Test BER performance with respect to distance.....	21

Table List

Table 1.	Impact of batch size	19
----------	----------------------------	----

Deep Learning-Based Optical Wireless Communication System Design

Sangshin Park

Department of Smart Robot Convergence and Application Engineering,
The Graduate School,
Pukyong National University

Abstract

Optical wireless communication (OWC) has been considered as a complementary or alternative technologies to radio frequency (RF) communications. OWC employs visible light emitted by light-emitting-diode (LED) or laser-diode (LD) to carry information which is known as visible light communication (VLC). Optical camera communication (OCC), a subsystem of OWC, uses LEDs as the transmitter and a camera or image sensor as a receiver. OCC can provide high signal-to-noise ratio (SNR) and noninterference communication even in outdoor environments.

This thesis investigates a deep learning (DL) framework for designing OCC systems where a receiver is realized with optical cameras capturing images of transmit LEDs. The optimum decoding strategy is formulated as the maximum a posteriori (MAP) estimation with a given received image. Due to the absence of analytical OCC channel models, it is challenging to derive the closed-form MAP detector. To address this issue, we employ a convolutional neural network (CNN) model at the OCC receiver. The proposed CNN approximates the optimum MAP detector that determines the most probable data symbols by observing an image of the OCC transmitter implemented by dot LED matrices. The supervised learning philosophy is adopted to train the CNN with labeled images. We collect training

samples in real-measurement scenarios including heterogeneous background noise and distance setups. As a consequent, the proposed CNN-based OCC receiver can be applied to arbitrary OCC scenarios without any channel state information. The effectiveness of our model is examined in the real-world OCC setup with Raspberry Pi cameras. The experimental results demonstrate that the proposed CNN architecture performs better than other DL models.



I. Introduction

Optical wireless communication (OWC) is a form of optical communication in which unguided infrared (IR), visible (VL) and ultraviolet (UV) bands to carry a signal. It is commonly used in short-range communication. With its powerful features such as low cost and high bandwidth, OWC can be a powerful alternative in contrast to existing wireless technologies. The emerging OWC has become popular in visible light communication (VLC) and free space optics (FSO). The use of OWC in mobile communication especially employed in 5G system which serves higher data rates.

VLC, a form of OWC using visible optical carrier produced by light emitting diode (LED) or laser diode (LD) has become a promising candidate for next generation wireless communication due to its advantages of long life expectancy and low power consumption. VLC system operates in the visible band (390-750nm). VLC takes advantage of light emitting diodes (LEDs) which can be pulsed at very high speeds without noticeable effect on human eye. VLC can be possibly used in a wide range of applications including wireless local area networks. On the other hand, terrestrial point-to-point OWC systems, also known as the FSO systems, operate at the near IR frequencies (750-1600 nm). These system typically use laser transmitters and

offer a cost-effective protocol. There has been growing interest on UV communication as a result of recent progress in solid state optical sources operating within solar blind UV spectrum (200-280nm). Such designs are particularly useful for outdoor line-of-sight (LOS) configurations to support low-power UV.

Optical camera communication (OCC) aims to deliver data from the light sources to the camera. OCC is different from VLC and Light Fidelity (Li-Fi) because various types of receivers are being used. With regard to the OCC, it is important to note that OCC-based systems are mainly targeted for a low-rate transmission, due to the reception-sampling rate, which is determined by the camera frame rate. Similar to VLC, OCC predominantly relies on the directional LOS transmission mode. In addition, OCC conveniently offers a MIMO capability by extracting the data from captured multiple light sources. Therefore, OCC can be considered as a subset of VLC.

VLC has been regarded as a promising solution to expensive radio frequency (RF) bandwidths. Visible light frequency bands can secure wide bandwidth in comparison to conventional RF communications without any authorization, thereby providing the cost-effective high data rate services [1]. The performance of the VLC can be further improved by employing optical

cameras at receivers [2], [3] where transmitted information is detected based on received images. This OCC system successfully distinguishes multiple lighting sources in the pixel domain, facilitating spatial multiplexing strategies via an array of transmit LEDs.

The optimum OCC decoding policy requires a novel signal processing strategy handling two-dimensional received images. The OCC channels, i.e., the propagation environments from transmit LEDs to receive cameras, have not been adequately studied in the literature. There have been several efforts on the analytical OCC channel models [4], [5], but they are typically confined to ideal scenarios in the absence of background noise and linear channel transfer functions. These are obviously far from practical realizations of the OCC systems suffering from various interruptions incurred by nonlinearity of LED and cameras, LED-camera misalignments, and ambient light interference. The lack of the exact channel state information (CSI) poses fundamental challenges in developing the maximum a posterior (MAP) detector for the OCC systems. For this reason, existing studies rely on simple image processing mechanisms [3] or additional LEDs [6]. These can only be suitable for stationary cases where both transmitter and receiver are fixed. To cope with time-varying OCC scenarios such as vehicular networks [3], it is essential to estimate CSI periodically. A recent work [1] has

investigated pilot-aided channel training techniques in the OCC system. Conventional multiple antenna communication strategies in the RF band are employed in designing the OCC receiver. The experimental results validate the feasibility of the CSI-assisted OCC detection rules. However, it still requires periodic channel estimation processes in each coherence time, thereby resulting in prohibitive overheads in time-varying scenarios such as mobile receivers.

To tackle these difficulties, this paper develops a deep learning (DL) based OCC receiver that is robust to possible system impairment such as background noise, unknown CSI, and arbitrary distant scenarios. A data-driven optimization property of DL techniques are exploited to train unavailable OCC MAP detector. Recent progresses [7], [8] have verified the feasibility of the DL in optimizing VLC transceivers. These are, however, still limited to artificially generated training samples. It is still an open problem to check the viability of the DL techniques in real-world OCC receiver optimization tasks.

This thesis presents a convolutional neural network (CNN) approach to identify the MAP detection rule without any CSI knowledge. To this end, training data samples are carefully measured in practical OCC setups with various patterns of background noise and heterogeneous distant setups. The

structure of the proposed CNN is carefully optimized such that the average bit error rate (BER) performance is minimized. The demonstration of the trained CNN receiver is carried out by Jetson Nano embedded boards with GPU installed. The experimental results demonstrate that the proposed CNN can successfully decode the transmitted bitstreams conveyed by square dot LED matrix.



II. Data Generation

As illustrated in Fig. 1, our experimental setup includes a dot LED matrix (MAX7219) controlled by Arduino Uno and a receiver realized by NVIDIA Jetson Nano and Raspberry Pi camera v2. As an initial study, we consider an ideal scenario where the transmitter and receiver are perfectly synchronized. This setup is exploited for the data collection and demonstration of trained OCC receiver. The training of the OCC receiver is performed on a PC with an Intel i9-10900KF CPU, 32 GB of RAM, and RTX 3080 GPU.

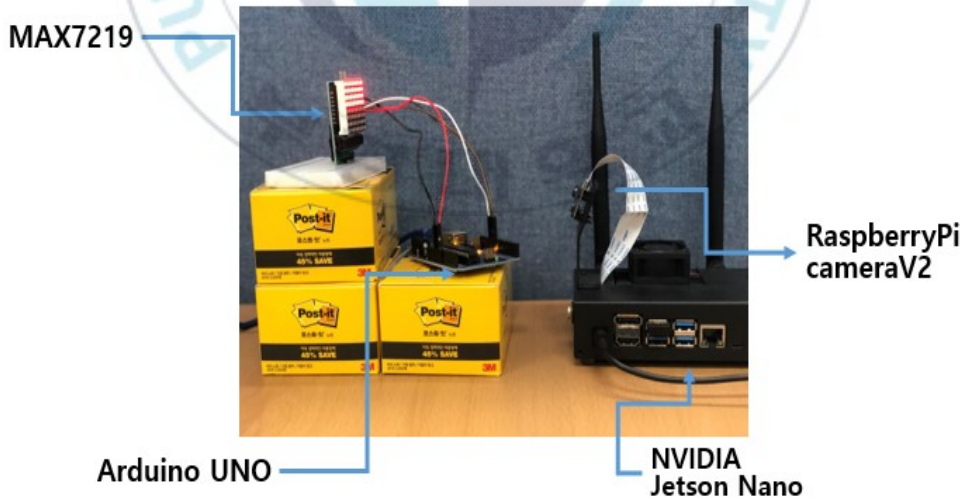


Fig. 1. Experimental environment.

2.1. Transmitter

The OCC transmitter is implanted by an 8-by-8 dot LED matrix shown in Fig. 2. On-off keying (OOK) modulation scheme is employed where the information bits are encoded by on-off patterns of multiple LEDs. A group of LEDs, in particular, neighboring 4-by-4 LED submatrix, conveys a single bit jointly. This results in the data rate of 4 bits per channel use. However, due to the data-driven optimization property of the DL techniques, the proposed approach can be readily extended to arbitrary number of the data rate. For the rest of this paper, it is assumed that total N bits are transmitted for each channel use.

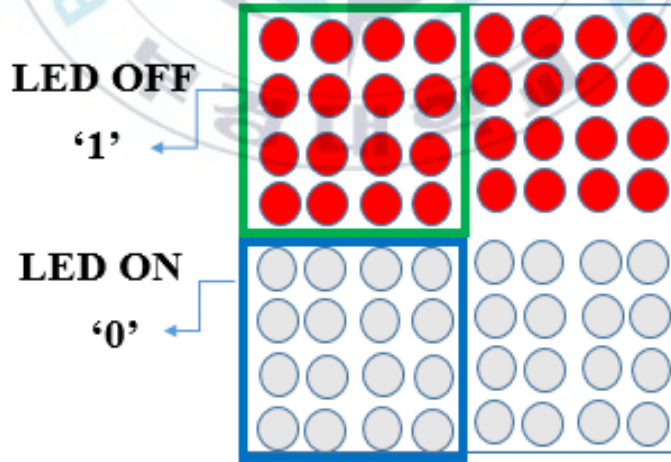
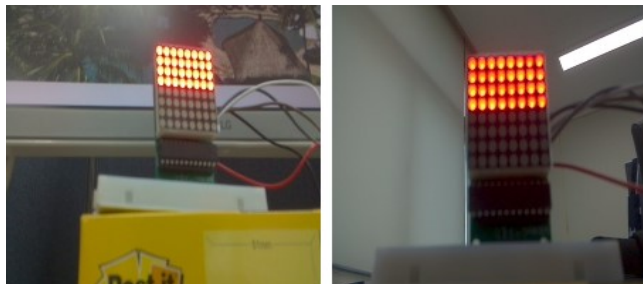


Fig. 2. Transmit LED matrix and OOK modulation strategy.

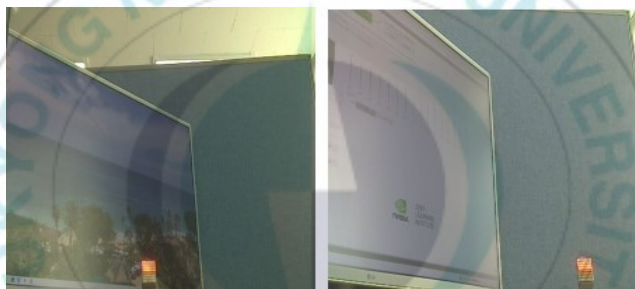
2.2. Receiver

For implementing CNN-based image processing, the receiver is realized by Jetson Nano boards having multi-core GPU. It is equipped with a single Raspberry Pi camera capturing images of the transmit LED matrix. The resulting images are resized into 28-by-28 size, and then are processed by the proposed CNN-based OCC receiver. The performance of DL models highly depends on training samples. To reflect practical OCC systems, we collect numerous instances of received images in various environments. Several examples are shown in Fig. 3. Training samples are collected at eleven different configurations according to the distance between transmitter and receiver: 12 cm, 20 cm, 25 cm, 30 cm, 35 cm, 40 cm, 50 cm, 70 cm, 100 cm, 130 cm, and 160 cm. At each setup, 16,000 images are collected, resulting in total 176,000 samples. All training data is labeled by its ground-truth bitstreams. Various OCC channel impairment features are included in the dataset. In particular, the transmit LED matrix is randomly rotated in a certain location (Fig. 3(a)) so that the proposed DL receiver can capture the effect of arbitrary misalignment between the transmitter and receiver. Also, numerous patterns of the background images are included as shown in Fig. 3(b) to develop the robust OCC receiver. Besides, the location of the transmitter randomly changes in each sample (Fig. 3(c)). These channel impairments can

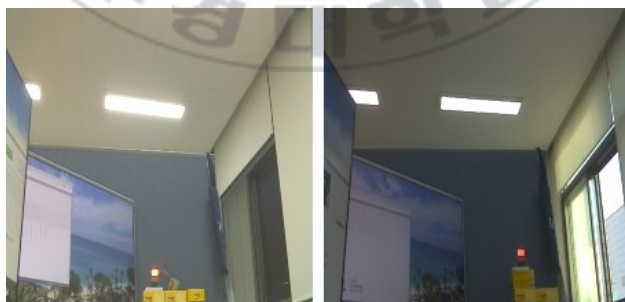
identify the optimum detection rule that is robust to arbitrary OCC scenarios without exact CSI knowledge.



(a) Samples with random misalignment angles (12cm)



(b) Samples with random background patterns (70cm)



(c) Samples with randomly located LEDs (160cm)

Fig. 3. Sample images of training dataset.

III. Proposed CNN-based OCC Receiver

The received RGB image is represented by a matrix $\mathbf{Y} \in \mathbb{R}^{28 \times 28 \times 3}$ where the last dimension stands for RGB color domain. Let $\mathbf{x} = (x_1, \dots, x_N)^T \in \{0,1\}^N$ be a set of N transmitted bits. These are encoded by the OOK modulation illustrated in Fig. 2 and are conveyed jointly. Then, the received image through nonlinear OCC channel transfer function $h(\cdot)$, which includes any interruptions from the transmit LEDs to the received image, can be written by

$$\mathbf{Y} = h(\mathbf{x}). \quad (1)$$

The optimum MAP receiver over such a OCC channel requires the exact and analytical formula of $h(\cdot)$, which is, in general, not available. To handle this difficult, we identify the robust OCC receiver that does not need any CSI knowledge. With practical bit interleaving operations, the transmitted bits x_1, \dots, x_N becomes independent. Therefore, a posterior probability of the received image \mathbf{Y} can be expressed by

$$p(\mathbf{x}|\mathbf{Y}) = p(x_1|\mathbf{Y}) \times \dots \times p(x_N|\mathbf{Y}). \quad (2)$$

Here, each marginal a posterior $p(x_n|\mathbf{Y})$ is computed as

$$p(x_n|\mathbf{Y}) = \mathbb{E}_h[p(x_n|\mathbf{Y}, h)], \quad (3)$$

where $\mathbb{E}_X[\cdot]$ stands for the expectation operation over a random variable

X . It is not difficult to determine $p(x_n|\mathbf{Y})$ when the OCC channel mapping $h(\cdot)$ and its distribution of are available. However, accurate channel models are typically not available in the OCC system due to the nonlinearity of LED and camera as well as the highly stochastic nature of image noise. Consequently, there is no analytical formula for a posterior $p(x_n|\mathbf{Y})$. To address this issue, we approximate the computation of a posterior by using CNN. It has been stated in the universal approximation theorem [9] that any continuous-valued functions can be successfully modeled by finite-layer neural networks with arbitrary small approximation error. Since a posterior (2) can be viewed as continuous-valued mapping of the received image, this can also be accurately modeled by carefully constructed deep neural networks.

Among various candidates, we choose the CNN architecture which has been known to be powerful for handling image dataset. Let $\mathbf{p} = \mathcal{F}_\theta(\mathbf{Y})$ be a CNN with a trainable parameter set θ . It produces N -dimensional a posterior probability vector $\mathbf{p} = [p(x_1 = 1|\mathbf{Y}), \dots, p(x_N = 1|\mathbf{Y})]^T \in \mathbb{R}^N$, i.e., each element of the CNN output represents the marginal probability $p(x_n = 1|\mathbf{Y})$ in (3). This can be regarded as discriminative models for binary classification tasks [10]. To yield a feasible marginal probability vector, the sigmoid function $\sigma(z) = 1/(1 + e^{-z})$ is employed as the output activation function of the CNN, whereas the rectified linear unit (ReLU) activation $\text{ReLU}(z) =$

$\max\{0, z\}$ is adopted at hidden layers. As a result, the prediction of the CNN for each transmitted bit x_n is written by $p(x_n = 1|\mathbf{Y}) = [\mathcal{F}_\theta(\mathbf{Y})]_n$, where $[\mathbf{z}]_n$ stands for the n -th element of a vector \mathbf{z} .

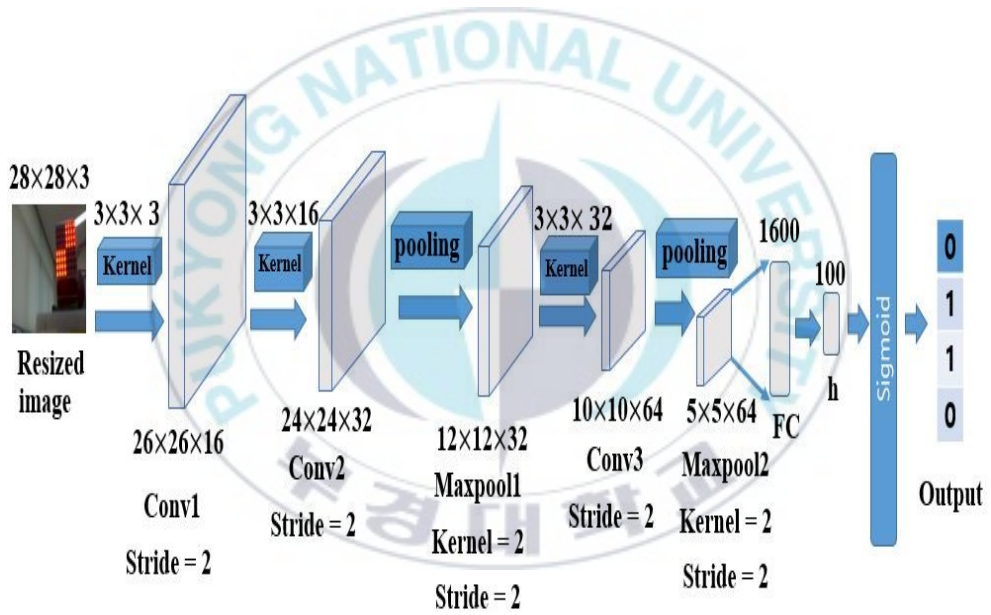


Fig. 4. Proposed CNN structure.

Fig. 4 depicts the proposed CNN architecture consisting of several convolutional layers followed by fully-connected (FC) layers. As mentioned, the input to the CNN is a resized image of size $28 \times 28 \times 3$ having RGB color channels. The first convolutional layer handles the input image with 16 kernels of size $3 \times 3 \times 3$. The stride is set to 2. This results in a latent feature of size $26 \times 26 \times 16$. The ReLU activation is applied to all hidden layers. At the second convolution layer, we use 32 kernels of size $3 \times 3 \times 16$, resulting in feature outputs of size $24 \times 24 \times 32$. The subsequent pooling layer performs down-sampling operation by taking the maximum value within 2×2 window. The third convolution layer consists of 64 $3 \times 3 \times 32$ kernels. After the second max-pooling operation, the resultant output becomes a 3D tensor of size $5 \times 5 \times 64$, which is fed to two FC layers. As discussed, the output activation is set to the sigmoid function to yield probability vector \mathbf{p} . The final decision for each transmitted bit can be obtained by the argmax operation.

The maximum likelihood (ML) estimate is adopted to optimize the CNN parameter set θ . With some manipulation, the negative log-likelihood of θ , denoted by $L(\theta)$, is expressed by

$$L(\theta) = \mathbb{E}_{(\mathbf{x}, \mathbf{Y})} \left[\sum_{n=1}^N l(x_n, [\mathcal{F}_\theta(\mathbf{Y})]_n) \right], \quad (4)$$

where the binary cross-entropy function is defined as

$$l(x_n, [\mathcal{F}_\theta(\mathbf{Y})]_n) = -x_n \log[\mathcal{F}_\theta(\mathbf{Y})]_n - (1 - x_n) \log(1 - [\mathcal{F}_\theta(\mathbf{Y})]_n).$$

In (4), the expectation is taken over all instances of (\mathbf{x}, \mathbf{Y}) including the stochastic OCC channel (1). The function $L(\theta)$ is exploited as a training loss of the proposed CNN. To minimize the loss function, we adopt the gradient-based training algorithm which iteratively updates the trainable parameter θ based on the gradient descent (GD) method as

$$\theta \leftarrow \theta - \eta \nabla L(\theta), \quad (5)$$

where $\eta > 0$ is the learning rate and ∇ denotes the gradient operator. For ease of implementation, the expectation term in (4) can be approximated by the sample mean over a training dataset $\mathcal{T} = \{(\mathbf{x}^{(i)}, \mathbf{Y}^{(i)}) | i = 1, \dots, T\}$ consisting of T samples of received images $\mathbf{Y}^{(i)}$ and their corresponding labels, i.e., the transmitted bitstream $\mathbf{x}^{(i)}$. Consequently, the mini-batch stochastic GD (SGD) updates of (5) can be written by

$$\theta \leftarrow \theta - \eta \frac{1}{S} \sum_{(\mathbf{x}, \mathbf{Y}) \in \mathcal{S}} \sum_{n=1}^N \nabla l(x_n, [\mathcal{F}_\theta(\mathbf{Y})]_n), \quad (6)$$

where $\mathcal{S} \subset \mathcal{T}$ with $|\mathcal{S}| = S$ accounts for a mini-batch set with randomly chosen S training samples.

IV. Simulation Results

This section assesses the performance of the proposed DL-based VLC systems. Among total 16,000 samples, 5280 and 5280 images are used for the training and validation, respectively, whereas the final evaluation of the trained CNN is conducted on the remaining 5440 test samples. The Adam optimizer [11] is adopted as the training algorithm. The batch normalization technique [12] is applied to all hidden layers.

The performance of DL models highly depends on choices of hyper-parameters, e.g., learning rate, network structure, and batch size. To validate the proposed CNN model, Fig. 5. shows the convergence behavior of the CNN training step for various learning rate values. In particular, we evaluate the validation BER performance with respect to the training epochs. It has been known that a large learning rate η leads to the fast convergence behavior, but at the same time, may incur a poor convergent point with unsatisfactory performance such as the saddle point and local minima. On the contrary, the training step requires intensive computations for a small η as it results in the slow convergence. Therefore, an appropriate learning rate should be carefully chosen based on the convergence behavior. From the figure, we can see that a large learning rate, e.g., $\eta = 0.01$ exhibits an to get

inefficient convergent point, whereas the result with $\eta = 0.0001$ suffer from the slow convergence speed. We thus choose an intermediate value $\eta = 0.001$ that achieves a good tradeoff between the validation BER performance and the learning speed.

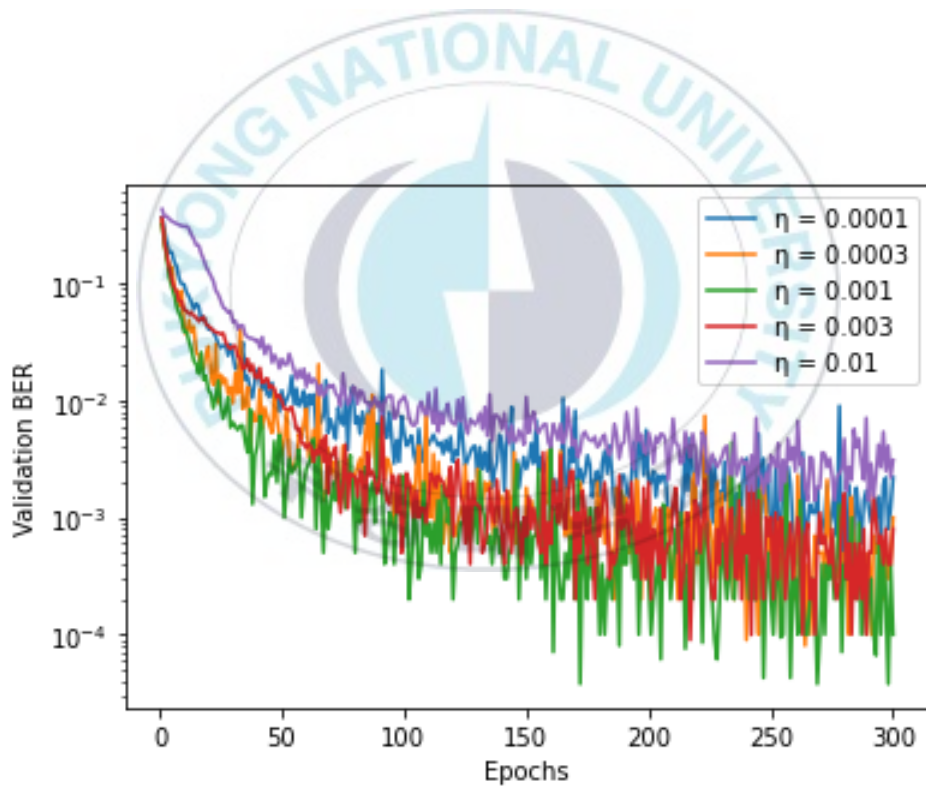


Fig. 5. Convergence behavior for various learning rate values.

Next, we examine the impact of the number of convolutional kernels in Fig. 6. Here, K represents for the number of kernels in the third convolutional layer. Based on Fig. 6, the performance of BER is generally better as K increases due to the increase of the expressive power of the CNN. However, employing a number of convolutional kernels requires high computational complexity in real-time decoding steps. Therefore, for designing practical OCC receivers with reasonable computing units, we need to choose a moderate K . We can see that the CNN with $K = 64$ achieves almost identical performance to that with $K = 512$. For this reason, the proposed CNN is constructed with 64 kernels at the third convolutional layer.

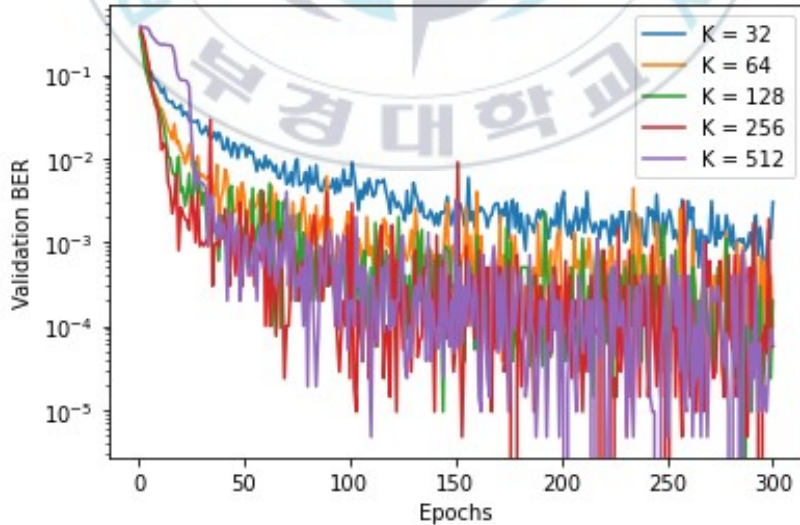


Fig. 6. Convergence behavior for various kernel configurations.

Fig. 7 illustrates the validation BER for various CNN configurations with the different number of convolutional layers C . Increasing C leads to a deepen structure with a higher model complexity, possibly posing the over-fitting problem. On the contrary, it would be insufficient to process received images with shallow CNNs, i.e., the case of a small C . Such an under-fitting and over-fitting tradeoff can be observed from the figure. The validation BER first increases as C grows and achieves the best performance at $C = 3$. However, the validation BER decreases after $C = 3$ due to the over-fitting issue. Therefore, an efficient choice is given by $C = 3$.

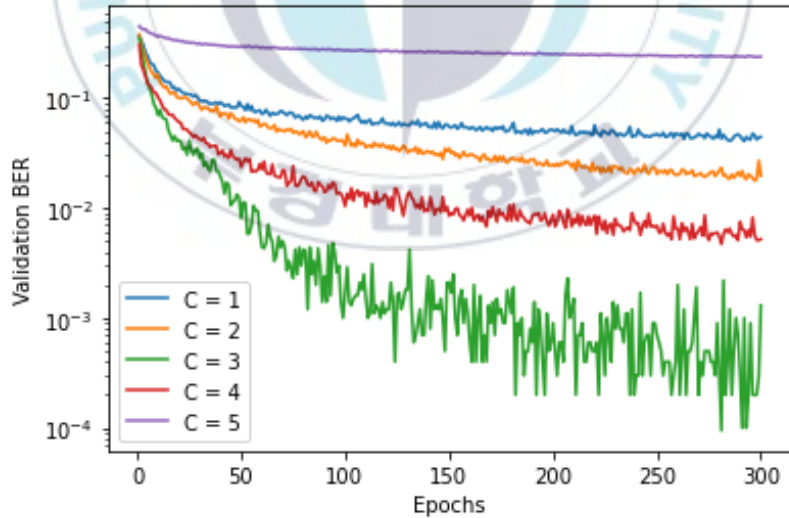


Fig. 7. Convergence behavior for various layer setups.

The impact of the batch size S to the BER performance is presented in Table I. It is shown that the validation BER performance is gradually improved as the batch size increases. We thus choose $S = 128$, which is the maximum allowable value for our training environment.

Table 1. Impact of batch size

Batch size	Validation BER
16	6.16e-05
32	5.21e-05
64	3.79e-05
128	2.36e-05

Finally, Fig. 8 provides the BER performance evaluated over the test dataset by varying the distance between the transmitter and the receiver. As benchmark, we also plot the test BER performance of a fully-connected neural network (FNN) which only consists of five FC layers. For fair comparison, the number of trainable variables of the FNN is carefully set to similar with that of the proposed CNN model. Vectorized images become inputs of the FNN. The output dimensions of each layer are set to 32, 64, 128, 256, and 4, respectively. Both the CNN and FNN are trained with samples measured at all simulated distance setups, and their test BER is examined at a specific distance configuration. From the figure, we can observe that the proposed CNN outperforms the FNN baseline at all simulated distance setups. The test BER performance of both schemes generally increases as the communication distance gets larger. This is because the decoding of distant LED matrices is typically more difficult. Nevertheless, the proposed CNN exhibits the BER performance around 10^{-4} regardless of the distance, which is the practical BER value in real-world OCC systems. It is worth noting that the proposed CNN method does not require any optical channel statistics both in the training and implementation stages. As a result, the detection rule of the proposed DL-based OCC receiver is independent of the communication environment. Thus, we can conclude that this result validates the feasibility

of the proposed CNN detector in practical OCC systems.

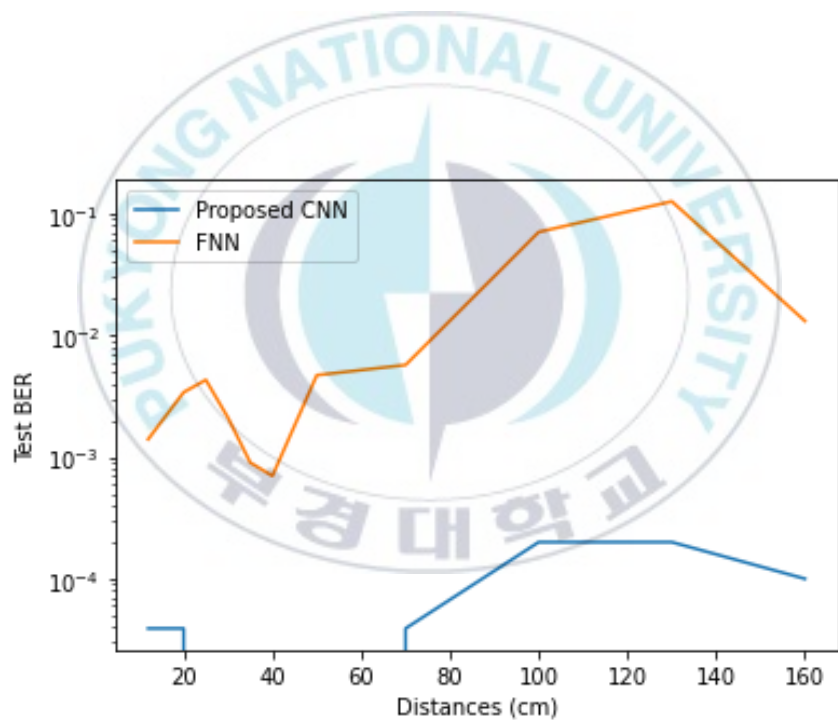


Fig. 8. Test BER performance with respect to distance.

V. Conclusion

This thesis investigates a DL approach for optimizing OCC receivers which robust to impairment of optical channels. To this end, a CNN is constructed which infers transmitted data bits from received images capturing dot LED matrices. Training data samples are carefully collected such that the CNN can learn statistical properties of arbitrary channel corruptions including random distance, background, and misalignment. A supervised training algorithm is presented which optimizes the CNN to mimic the optimal MAP detector without the CSI. The effectiveness of the proposed CNN-based OCC receiver is examined real-world experimental environments. It is revealed that the DL techniques can infer arbitrary optical channel impairments in a data-driven manner and achieves practical ranges of BER performance. It would be interesting future work to reduce the real-time computational complexity of DL-enabled OCC receivers. A possible solution is to design a light-weight CNN based on the knowledge distillation technique [13].

References

- [1] Y. Ohira, T. Yendo, S. Arai, and T. Yamazato, "High performance demodulation method with less complexity for image-sensor communication," *Opt. Express*, vol. 27, no. 15, Jul. 2019.
- [2] T. Nguyen, A. Islam, T. Yamazato, and Y. M. Jang, "Technical issues on IEEE 802.15.7m image sensor communication standardization," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 213 – 218, Feb. 2018.
- [3] T. Yamazato, I. Takai, H. Okada, T. Fujii, T. Yendo, S. Arai, M. Andoh, T. Harada, K. Yasutomi, K. Kagawa, and S. Kawahito, "Image-sensor-based visible light communication for automotive applications," *IEEE Commun. Mag.*, Vol. 52, pp. 88 – 97, Jul. 2014.
- [4] J. Perez-Ramirez and D. K. Borah, "A single-input multiple-output optical system for mobile communication: modeling and validation," *IEEE Photonics Tech. Lett.*, vol. 26, no. 4, pp. 368–371, Feb. 2014.
- [5] Y. Sun, D. K. Borah, and E. Curry, "Optimal symbol set selection in GSSK visible light wireless communication systems," *IEEE Photonics Tech. Lett.*, vol. 28, no. 3, pp. 303-306, Feb. 2016.
- [6] W. A. Cahyadi, Y. H. Kim, Y. H. Chung and C. Ahn, "Mobile phone camera-based indoor visible light communications with rotation compensation," *IEEE Photonics J*, vol. 8, no. 2, pp. 1-8, Apr. 2016.
- [7] 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, Mar. 2018.
- [8] H. Lee, S. H. Lee, T. Q. S. Quek, and I. Lee, "Deep learning framework for wireless systems: Applications to optical wireless communications," *IEEE Commun. Mag.*, Vol. 57, no. 3, pp. 35-41, Mar. 2019.

- [9] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.
- [10] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [11] D. Kingma and J. Ba, “Adam: a method for stochastic optimization,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.
- [12] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariance shift,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 448–456, July 2015.
- [13] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Dec. 2014