Bubble-Wave-Mitigation Algorithm and Transformer-Based Neural Network Demodulator for Water-Air Optical Camera Communications

Maolin Li, Peng Ling, Shangsheng Wen, Xiandong Chen, and Fei Wen

***Abstract*—Optical camera communication (OCC) has been widely employed in various applications as a flexible and cost-effective means of communication both on land and underwater. However, the performance of the OCC system through the water-air interface has not been thoroughly investigated. In this article, we explore the performance of the OCC system in a water-air environment and propose a bubble-wave-mitigation algorithm to pre-process the captured frames of received video. Moreover, we propose a transformer-based neural network to demodulate the transmitted signal, mitigating the deterioration in transmission performance caused by inter-symbol interference (ISI). The experimental results demonstrate that a robust transmission can be achieved in the water-air environment by applying our proposed algorithms and neural network demodulator.**

***Index Terms*—Water-air optical camera communications, waves, bubbles, rolling shutter effect (RSE), RGB-LED, underwater optical camera communication (UOCC), deep neural network.**

I. Introduction

UNDERWATER wireless communication (UWC) has become a popular subject of discussion due to the rise of human activities underwater, such as underwater exploration, rescue missions, pollution monitoring, and underwater salvage [1]. Acoustic wave-based UWC technologies are currently the most widely used due to their ability to facilitate long-range transmission underwater, albeit with high transmission losses and latency. Moreover, most energy of acoustic wave will be reflected off the surface of the water and it will suffer unacceptable attenuation in the air. Radio frequency (RF), on the other hand, provides high transmission rates and a larger bandwidth, but its high attenuation rate in underwater environments limits its transmission range [2]. Optical wireless communication (OWC) presents a promising solution for underwater data transmission as it offers improved security, lower latency, and faster transmission speeds. Additionally, it is a potential candidate for water-air communication systems. Compared to RF and acoustic signals, the optical signal is preferred due to its lower loss through the water-air interface and limited propagation attenuation for both air and water links [3]. Utilizing floating buoy stations with acoustic or radio transceivers for relay transmission is a viable option. However, this approach requires early deployment and incurs higher costs. Moreover, it may prove impractical in high-risk application scenarios such as combat situations and lacks the necessary speed for addressing emerging events [4]. In contrast, employing light-based signal transmission offers a comparatively lower cost and easier deployment. When deployed in a heterogeneous edge cluster, [5], [6] can help reduce the network latencies. [7], [8], [9] provide solutions for event-detection and other tasks in the resource-constrained underwater environment.

Despite its potential, there are still challenges that impede the widespread implementation of water-air OWC. Firstly, ocean turbulence causes fluctuations in power at the receiver and produces scintillation effects. Secondly, apart from blue or green wavelengths, other wavelengths of light experience relatively high attenuation underwater. The presence of phytoplankton and detritus in the water causes scattering, further impacting the transmission of light signals [10]. Consequently, most current research on Optical Wireless Communication (OWC) focuses on blue or green light sources. Thirdly, the presence of air bubbles in water results in reflection and refraction, leading to significant performance degradation in optical signal propagation [11]. Additionally,the optical signal passing through waves is prone to refraction, which can distort or blur images captured by the receiver. The interface between water and air is highly complex and has garnered significant interest, and there is some research on the topic. The water-air OWC remains an open problem for researchers to explore.

Optical camera communication (OCC) is a sub field of OWC that utilizes image sensors as receivers instead of photo detectors (PD) in water-air communication systems. While its transmission bandwidth may be lower than PD-based OWC, OCC offers a more flexible and convenient means of achieving optical communication by utilizing widely distributed embedded cameras in electronic devices [12]. In a Photodiode (PD)-based Optical Wireless Communication (OWC) system, the receiver relies on detecting the received voltage level to decode the data. However,

Manuscript received 26 July 2023; accepted 3 August 2023. Date of publication 7 August 2023; date of current version 18 August 2023. This work was supported by the National Undergraduate Innovation and Entrepreneurship Training Program under Grant 202210561186. *(Corresponding author: Shangsheng Wen.)*

Maolin Li and Peng Ling are with the School of Automation Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510640, China (e-mail: 202030461169@mail.scut.edu.cn; 201930133152@mail.scut.edu.cn).

Shangsheng Wen and Xiandong Chen are with the School of Materials Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510640, China (e-mail: shshwen@scut.edu.cn; cxdscut@163.com).

Fei Wen is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: fei8wen@gmail.com).

Digital Object Identifier 10.1109/JPHOT.2023.3302690

when implementing this system in an underwater environment, decoding using photo detectors becomes challenging [13]. This difficulty arises from the low received power and high cross-channel interference caused by the absorption and scattering of light waves in water, as mentioned previously. In such cases, utilizing image-based camera receivers proves advantageous compared to using photo detectors, as the decrease in visibility is not significant [13], [14]. Additionally, when decoding data using a camera-based receiver, effective cross-channel interference cancellation can be achieved through image processing. The spatial separation feature of the camera also helps in distinguishing the presence of multiple light sources and mitigates environmental light noise and other sources of interference [15]. In OCC systems, LED is commonly used because visible light with a limited beam angle offers a good balance between penetration and coverage. In contrast, using highly directional laser beams presents practical challenges in terms of directivity control for high-speed links, especially in the presence of oceanic turbulence. Precisely targeting the receiver becomes difficult due to the movement of the receiver in the water [16]. Typically, the camera's field of view (FOV) is greater than 50 degrees, enabling image sensor-based technology to achieve reliable optical signal transmission without strict alignment. This increases the flexibility of the communication system and improves the quality of optical wireless communication in complex scenarios, such as outdoor and underwater environments. Image sensors have two shutter modes: global shutter mode and rolling shutter mode. A complementary metal-oxide semiconductor (CMOS) image sensor with rolling shutter effect (RSE) is commonly adopted in communication because its sampling rate is higher than the frame rate. This enables the communication system to receive multiple bits within one frame, resulting in a higher transmission rate [17], [18], [19]. The transmission distance of RSE-based OCC typically ranges from tens of centimeters to a few meters. OCC is now being employed in underwater environments.

The availability of robust LED lighting systems in shallow or deepwater applications and underwater vehicles, in addition to camera systems that are frequently integrated into diverse marine devices, including drones, scuba diving equipment, and camera sensor nodes for monitoring purposes, has contributed to the emergence of optical camera communication (OCC) systems as viable alternatives [16], [20]. Although OCC is primarily suitable for medium and short-range applications, it suffices to transmit data from a submerged diver or a launcher to a ship or boat, which falls within the short to medium range transmission for most UWC applications [14]. Researchers in [14] have designed and implemented an OCC system for underwater wireless communication purposes. Furthermore, in [11], rolling shutter-based OCC was investigated in a strong bubble underwater environment, and a data rate of 7.2Kbit/s was achieved. OCC provides a flexible and cost-effective approach for implementing underwater optical communication (UWOC) systems, harnessing the advantages of both lighting and communication. The water-air optical camera framework explored in this research holds potential for diverse applications, including navigation lighthouses, navigation landmarks, and landscape lighting. This framework can be employed by leveraging existing lighting devices or deploying new hardware equipment underwater. For instance, in the context of creative lake landscape lighting, existing RGB headlights installed at the lake's bottom can be utilized. Users only need to employ their phone cameras to receive the transmitted information, making this a low-cost communication method that facilitates future commercial applications. However, rolling shutter-based OCC has not been well-researched in water-air communication systems.

Currently, RGB-LED-based OCC systems are used with different modulations, such as color-shift keying modulation, to improve data rate performance [21]. However, conventional decoding algorithms do not perform well at higher speeds due to inter-symbol interference (ISI). This can be attributed to the operational behavior of rolling shutter-based cameras, which do not wait for the previous row to complete exposure before starting the next row. As a result, the exposure times of adjacent rows overlap [22]. Thus, more effective and robust decoding algorithms are necessary to address these issues. In recent years, deep neural networks have demonstrated potential in OCC systems due to their powerful recognition and classification abilities. In [23], a Convolution Neural Network (CNN) has been introduced as a classifier to recognize the LED-Identity(LED-ID). In [24], an Long Short Term Memory(LSTM)-based equalizer has been proposed to decode. In 2017, the transformer model was introduced and has been successful in natural language processing (NLP) [25], computer vision (CV) [26], and other fields due to its ability to deal with time-sequence problems and extract contextual information. However, researchers have not paid much attention to its potential use in the OCC field. The main contributions in this article are summarized as follows:

- We are the first to investigate the rolling shutter effect (RSE)-based OCC system in the water-air environment, and we also use RGB-LED as a trasmitter to improve data rate.
- We propose an image processing algorithm to reduce adverse effects of waves and bubbles on the water-air communication channel.Therefore, the data transmission is faster and more stable.
- We also introduce a transformer-based deep neural network to enhance the performance of OCC demodulation, which, to our knowledge, has not been explored in the field of OCC.

The article is structured as follows: Section II introduces the proposed system, and Section III illustrates the image pre-processing algorithm and the transformer-based model. In Section IV, we present the experiment results and analysis. Finally, Section V shows our conclusion.

II. Water-Air Optical Camera Communication System

Figs. 1 and 2 illustrates the architecture of the designed water-air optical camera communication system. Data packets are generated offline from a personal computer (PC) and loaded into the single port read-only memory (ROM) of a field-programmable gate array (FPGA) from Altera Cyclone IV series to control the RGB-LED. [27], [28] help improve the power efficiency of memory subsystem. Our experimental setup was implemented

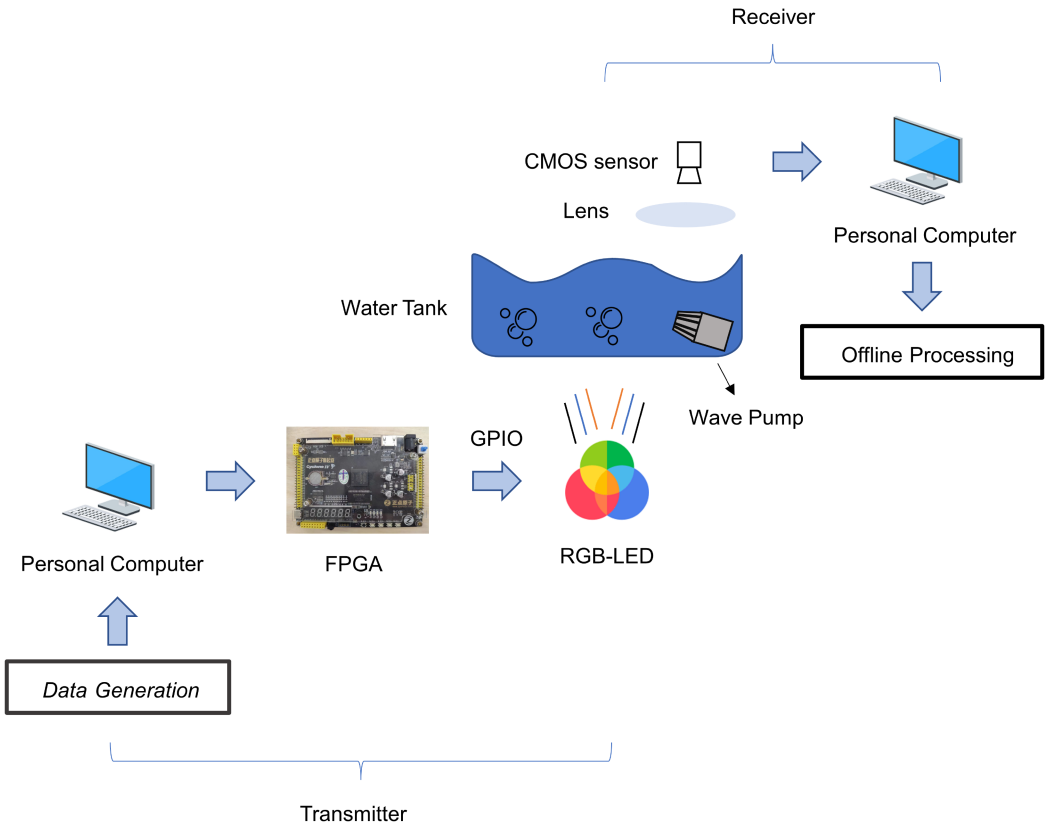


Fig. 1. Overview of water-air optical camera communication system.

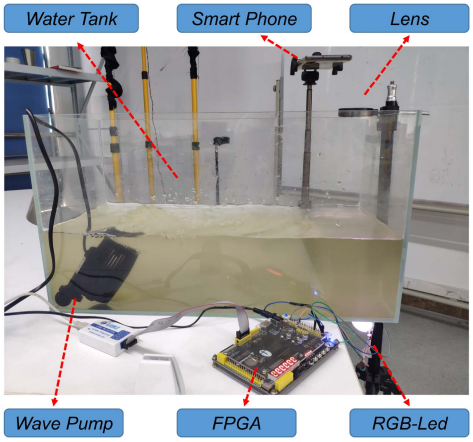


Fig. 2. Experiment physical setup.

TABLE I
SYSTEM SPECIFICATION

Parameters	Value
The resolution of the camera	1920*1080
The ISO of the camera	6400
Frame rate /fps	60
Exposure time /μs	250
Exposure compensation /EV	0
The focal length of lens /mm	49
Voltage of the LED	2.5
Power of the LED	10W
Lighting Angle of the LED	120
Lighting Angle of the RED-LED	620 nm
Lighting Angle of the GREEN-LED	520 nm
Lighting Angle of the BLUE-LED	465 nm

in a filled water tank with dimensions of 23 cm $\times$ 28 cm $\times$ 50 cm. The transmitted signal travels through a 20 cm water channel and a 40 cm atmospheric channel. A wave pump is used to simulate real water-air environments. At the receiver, a plano-convex lens is used in front of a smartphone (Huawei P20 Pro with a resolution of 1920 $\times$ 1080 and a frame rate of 60 fps) to condense the incoming light and increase the imaging area of the LED. The smartphone, which uses a CMOS-based optical camera, is set to video capture mode with a manual configuration of the exposure compensation of -4 EV, an ISO value of 6400, and an adjusted exposure time of 250 μs. Table I shows the key hardware specification.

The structure of the data packet is as shown in Fig. 3. Each data packet includes a 16-bit header and payload and is transmitted three times within (1/frame rate) second to ensure that the CMOS-based optical camera can capture it. In this work, we employed 8-CSK modulation, with on-off keying used for each channel, so that each symbol includes three bits.The general-purpose input/output (GPIO) pin of the FPGA outputs high and low voltage, corresponding to the data bits “1” and “0” from the data packets stored in ROM. The RED-LED, GREEN-LED,

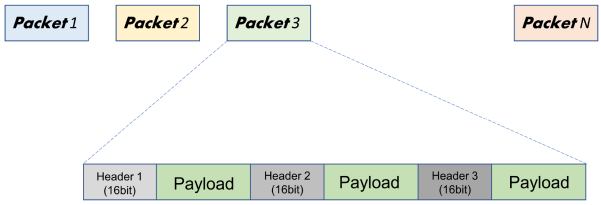
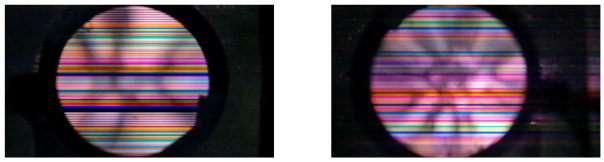


Fig. 3. Data packet structure of the transmitted signal.



(a) (b)

Fig. 4. (a) Captured frame distorted by the combination of bubbles and waves. (b) Captured frames severely blurred by intense waves.

and BLUE-LED are controlled by three GPIO pins of the FPGA, respectively.

In the receiver, we have used a 60fps camera of a smartphone to capture the images of the optical signal. The transmitted signal is captured through the camera using the rolling shutter effect. Utilizing an RSE-based camera, we can capture the changing states of a light source in a single image. The image is generated row by row, with each row of pixels being exposed and read out sequentially by a CMOS sensor. When optical signals transmit through the water-air interface, they are subject to severe distortion due to the refraction of waves and bubbles. The main challenge in underwater optical camera communication is the negative impact of bubbles, which causes optical attenuation. However, water-air optical communication faces a more challenging issue than underwater optical camera communication. In addition to the interference caused by bubbles in underwater environments, the most difficult problem is the changing water surface due to waves, resulting in image distortion and blurring. This situation not only increases the difficulty of locating the header but also exacerbates the interference between adjacent stripes, as illustrated in Fig. 4. In our proposed system, we use RGB LED, which amplifies this negative impact and makes it more difficult to decode using traditional threshold and sampling algorithms. To address this issue, we propose the bubble-wave-mitigation algorithm and a transformer-based neural network demodulator to process images in this environment. Firstly, our bubble wave elimination algorithm is applied for pre-processing of each frame to alleviate the negative impact of bubbles and waves and improve recognition accuracy of the header of the data packet. Next, a threshold method [29] is used for locating the header of the data packet. The header consists of two symbols, one for "brightness" and the other for "darkness," making it easier to recognize. Subsequently, we select a specific column L and extract sub-images consisting of 10 rows beginning from the location of header. The column matrix for each row is L-4 to L+5. All sub-images extracted from one frame are simultaneously fed into the network, as illustrated in Fig. 5. The trained network then classifies the sub-images and obtains the data packets in one frame. It is important to note that we feed the sub-images in the order of cutting, so the resulting output is also in the same order. After processing all frames, a bit error rate (BER) calculation is performed.

III. Design and Methodology

A. Bubble-Wave-Mitigation Algorithm

The proposed image pre-processing algorithm is aimed at improving the accuracy of signal decoding in the presence of bubbles and waves that can significantly affect the images captured by the CMOS-based optical camera. The shadow caused by bubbles and waves does not completely block the color information of the picture but it does affect the overall brightness distribution of the image. Therefore, our algorithm is based on row pixel sorting and brightness normalization. To avoid losing color information in some rows due to the presence of black shadows caused by waves and bubbles, the elements of each row in the gray scale image frame are sorted first, moving the shadows to the left and the columns with rich color information to the right as shown in Fig. 6(a). Then, the columns appearing on the right side of the sorted image from 1200–1900 are cropped as new images and subjected to further processing. The brightness of the cropped image is uneven, so selecting specific vertical columns for sampling may result in relatively large pixel value fluctuations in the column, making it difficult to determine an appropriate threshold for data decoding. To address this issue, we initially divide the image into three separate RGB channels for subsequent processing. This division is necessary because different wavelengths of light encounter varying levels of attenuation in water, which also helps to mitigate cross-channel interference. The average pixel values of each color channel's image are then computed, and the three channels of each pixel point in the captured image are divided by their corresponding average values. This step results in a brightness equalized image that effectively enhances image contrast. Next, we proceed by selecting a single column, and we determine the maximum value within that column for each color channel's image. These maximum values serve as reference pixel values. To normalize the image across all channels, we divide each pixel by its corresponding maximum value in each color channel. Finally, we employ a gamma correction method to compensate for the nonlinear distortion introduced by the hardware. This correction helps restore the original image characteristics. The resulting processed image, as depicted in Fig. 6(b), is obtained by replicating a single selected column multiple times.

B. Transformer-Based Neural Network Demodulator

The demodulation model's complete structure is illustrated in Fig. 5. We utilized the encoder architecture from transformer [25], which is well-known for sequence prediction. Recently researchers also utilize transformer-based models for image-related tasks [30], [31]. and we made some specific designs to cater to our practical requirements. Inspired by [32], the initial step in our transformer-based demodulation model

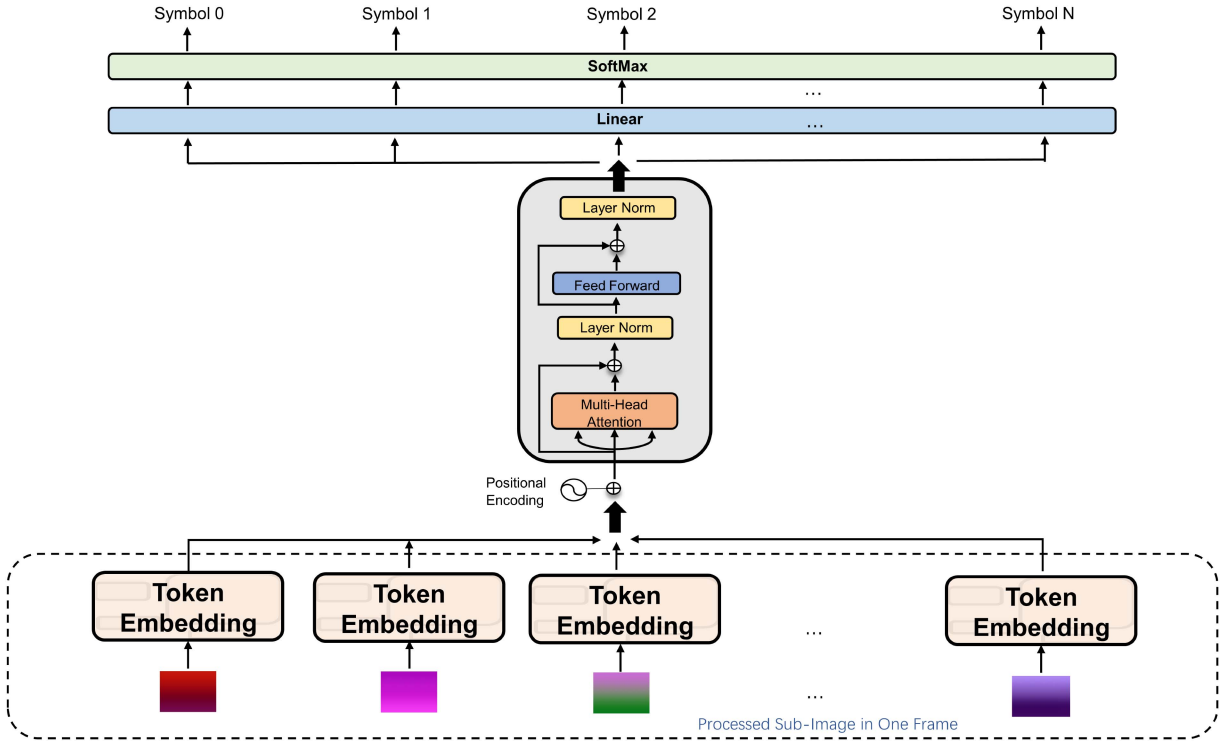


Fig. 5. Transformer-based neural network demodulator for optical signals.

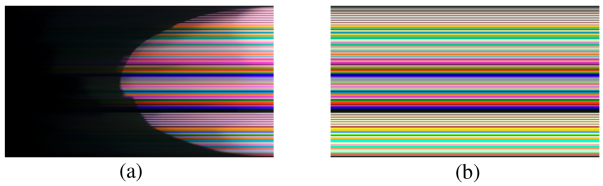


Fig. 6. (a) The sorted image start with Fig. 4(a). (b) The image processed by bubble-wave-mitigation algorithm. We select a column for processing and repeat it in the column dimensions to form an image to show.

is to tokenize every input image into a 512-dimensional vector which is called token-embedding. As demonstrated in Fig. 7, the primary component of the embedding network is the Squeeze-and-excitation-Residual(SE-Res) block, which combines the basic structure of Squeeze-and-excitation network (SE-net) [33] and Residual network (Res-net) [34]. It aims to determine the importance level of each channel in the feature map and assigns a weight value to each feature based on this importance level. This enables the neural network to concentrate on specific feature channels. The residual portion of the block makes it easier to optimize the network and achieve improved accuracy with a significantly increased depth. Additionally, Batch Normalization (BN) layers are used to stabilize and smooth the optimization problem and provide robustness to hyperparameter settings [35]. Finally, we apply average pooling to the multi-channel feature maps and flatten them into a 512-dimensional feature description vector as a token. After token-embedding, positional encoding vectors are added to these 512-dimensional vectors. We adopt the same positional encoding method as the original transformer

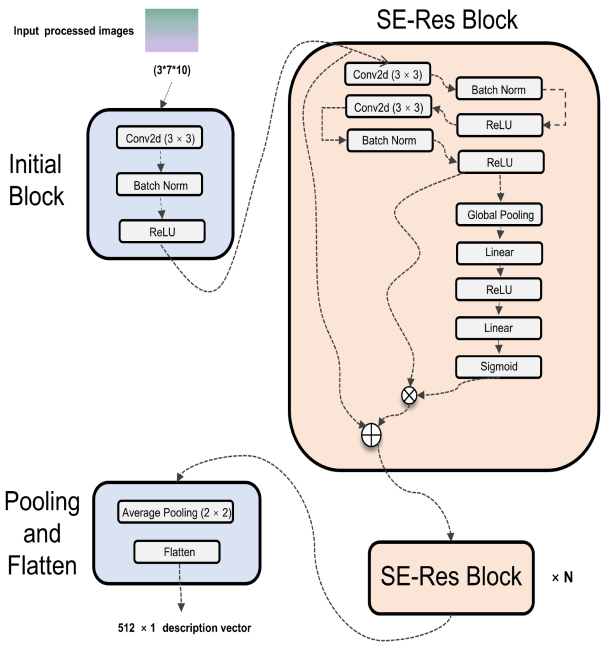


Fig. 7. Structure of token embedding layer which comprised of some SE-Res blocks.

architecture [25] to characterize the relative positional relationship of these symbols and it is expressed as:

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}}). \quad (1)$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}}). \quad (2)$$

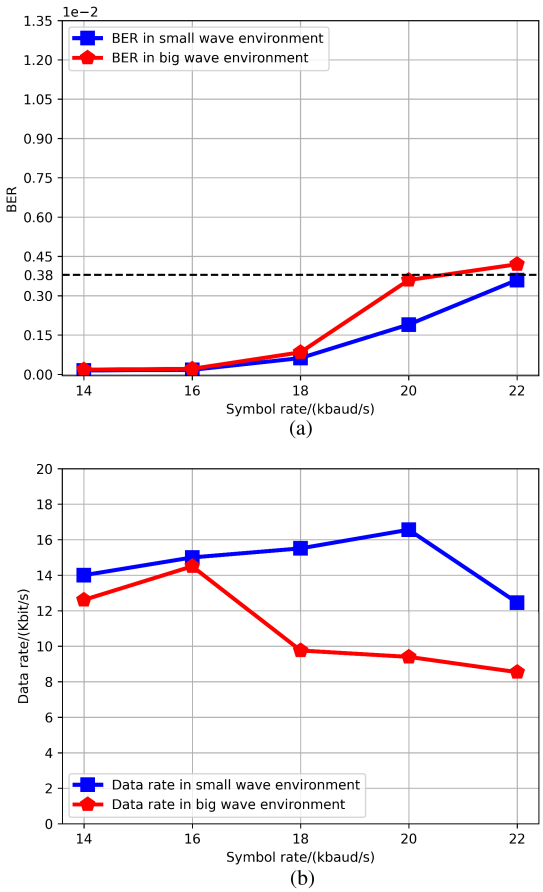


Fig. 8. (a) BER performance. (b) Data rate.

where pos is the position of a description vector in the entire input, for example, the pos of the first slice of the frame image is 0. $2i$ and $2i+1$represent dimensions, and the range of values for i is [0,..., 512/2], where 512 is the dimension of the description vector,. This step is crucial because inter-symbol interference (ISI) often occurs between adjacent symbols, and without positional information, it will be difficult for the model to focus on learning the relationships between adjacent symbols. The position encoding is added directly to each vector. Next, multi-head attention is introduced to learn the relationships between various vectors and extract higher-order features for classification. Its input and output are the same as that of Recurrent Neural Networks (RNN). However, RNN must process tokens sequentially, which is slow in runtime and is prone to the gradient vanishing or exploding problem [36], [37], [38]. In contrast, the multi-head attention does not use a sequential structure and can aggregate information from the entire input sequence, making it more suitable than RNN for integrating global and long-range information. Additionally, its powerful ability to process sequences can specifically mitigate ISI happening in the current sub-image demodulation, thereby reducing the classification error.

Then, we use residual connections and layer normalization. The input of the multi-head attention block is added to the output of the block to preserve the original input information and avoid network degradation. Layer normalization is used to reduce training time and improve the stability of network hidden states [39]. Next, a feedforward layer is applied, followed by residual connections and layer normalization again. Finally, we utilize three linear layers with Rectified Linear Unit (ReLU) as the activation function. ReLU's sparsity property enables the model to effectively extract relevant features and fit the training data. Additionally, ReLU has low computational complexity, as activation values can be obtained with a single threshold, thereby improving training and testing time. This is followed by another linear layer with SoftMax activation to classify the label of each output vector, which corresponds to symbols in the transmitted data packet.

The goal of our training is to minimize the cross-entropy loss. To avoid memory effects in the neural network, we used two videos capturing randomly generated messages, which are generated separately for training and testing. The frames in the training video are preprocessed using our bubble-wave-mitigation algorithm. Next, the sub-images that represent the corresponding symbols in each frame will be treated as a sequence and fed into the transformer-based network for training. We set the number of training epochs to 30 and used Adaptive Momentum Estimation (Adam) as the optimizer. We also optimized some hyperparameters such as the number of SE-Res blocks and kernel size, and the number of multi-head attention blocks to achieve better performance while keeping the network from becoming too deep.

IV. Experiment Result and Analysis

A. Data Rate and BER Result

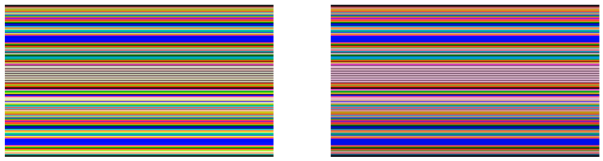
We initially investigated the performance of our proposed algorithms by transmitting symbols at rates of 14 kBaud, 16 kBaud, 18 kBaud, 20 kBaud, and 22 kBaud. We conducted experiments in both small waves, which cause fewer shadows, and big waves, which cause more severe image damage. The horizontal distance between the center of the wave pump and the phone camera is 45 cm when generating small waves. This distance is reduced to 25 cm when generating large waves. As for the bubbles, they are by-products of the wave generated by the wave pump and are randomly formed during vibration and spraying. Most of them float on the water's surface, with sizes ranging from 0.3 to 0.8 cm. Table II summarizes the hyperparameters of the network we chose for the experiments. Fig. 8(a) shows the BER performance at different symbol rates. For all symbol transmission rates in the small waves environment, the BER was below $3.8 * 10^{-3}$, meeting the forward error correction (FEC) requirement. However, at a symbol rate of 22 k in the big waves environment, the BER could not meet the FEC requirement. Furthermore, Fig. 8(b) reveals that a higher symbol transmission rate may not always lead to a higher data rate due to severe image distortion in wave environments, which increases the difficulty of decoding. When the data rate exceeds 20 kbaud/s, even in environments with small waves, packet loss and a decrease in rate can occur. Furthermore, in environments with big waves, the data rate decrease begins at 16 kbaud/s due to severe image blurring,which exacerbates inter-symbol interference(ISI). Overall, our proposed preprocessing algorithm and

TABLE II
NETWORK PARAMETERS

Parameters	Value
Number of SE-Res blocks	4
Kernal size	3×3
Number of Transformer-based blocks	3
Number of attention-heads	4
Number of nodes of Feed Forward Layer	1024
Training weight decay	5e-4
Training drop out	0.1

TABLE III
COMPARISON BETWEEN DIFFERENT SYSTEMS

Data rate	Distance	Ref	Methods for mitigating bubbles or waves	Water type
16.56 Kbit/s	1 m	Our scheme	Both	lake water
7.2 Kbit/s	1.5 m	[11]	Bubbles	pure water
100 b/s	1 m	[14]	None	pure water
750 b/s	1 m	[13]	None	pure water



(a) (b)

Fig. 9. (a) The image processed with our algorithm. (b) The image processed with algorithm in [11].

demodulation framework demonstrated good performance for OCC systems in water-air environments, maintaining a high data rate and low BER despite interference from big waves. This suggests that our approach can effectively address the challenges posed by water-air communication channels.

B. Performance Comparison

We conducted a performance comparison between our Bubble-Wave-Mitigation Algorithm and the De-Bubble Algorithm proposed in [4]. In their work, the researchers utilized an RSE-based CMOS image sensor as a receiver in underwater OCC systems and proposed a de-bubble algorithm to counter bubble degradation. However, they did not test their algorithm using RGB-LED or at a higher transmitted symbol rate. As depicted in Fig. 9, a comparison of the preprocessed images generated by our algorithm and theirs reveals that our images exhibit better quality, more uniform brightness, less color deviation, and a more recognizable header. This is due to the fact that we meticulously observed the images disrupted by bubbles and waves and implemented suitable preprocessing steps. Regarding the data rate, their system achieved 7.2 Kbit/s, whereas ours achieved 16.56 Kbit/s. A comparison of the data rate with other underwater OCC systems is shown in Table III, indicating that our proposed scheme achieves a faster data rate. Furthermore, our scheme is less susceptible to interference, and we believe that our water-air optical camera framework can be applied to creative applications such as navigation lighthouses, navigation landmarks, and landscape lighting. Although our algorithm requires more computing resources, we believe that this can be addressed by utilizing special chips and circuits.

C. Ablation Study

To validate the capabilities of our proposed algorithm in adaptively addressing image distortion caused by the adverse effects of water on various light wavelengths, we conducted a comparative analysis. We examined the impact of applying normalization and gamma correction steps within the bubble-wave-mitigation algorithm on the final decoding result under both small wave and big wave conditions. The results as Fig. 10 shows revealed that the exclusion of normalization and gamma correction led to an unacceptable increase in the error rate. In contrast, experiments employing the complete algorithm demonstrated superior outcomes, providing compelling evidence that our algorithm effectively resolves issues related to shadowing, occlusion caused by wave and bubbles and the light attenuation in underwater environments. However, the data rate did not experience a significant decrease. We will provide a more detailed explanation for this phenomenon in subsequent ablation experimental groups. Next, to demonstrate the effectiveness and advanced performance of our transformer-based neural network demodulator, we conducted experimental research to evaluate the deep feature extraction ability of the transformer framework and the performance improvement of token embedding using SE-Res blocks. Two sets of experiments were designed, one using our proposed network demodulator, and the other using token embedding based on SE-Res block in our demodulator but without transformer-based feature extraction. Each group of experiments was conducted separately in small wave and large wave environments. In the second set of experiments, after token embedding, we used linear layers and SoftMax for classification, without the same operations as traditional transformers such as position encoder and multi-head attention. Figs. 11 and 12 show the BER performance and data rate at symbol rates ranging from 14 kbaud/s to 22 kbaud/s. The reduction in the bit error rate is significant, but there is only a moderate improvement

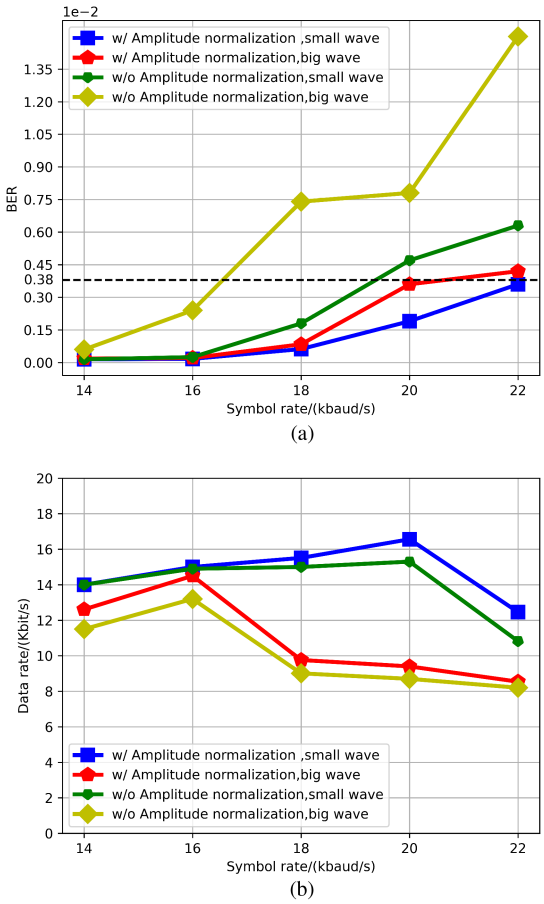


Fig. 10. (a) BER performance comparison. (b) Data rate comparison.

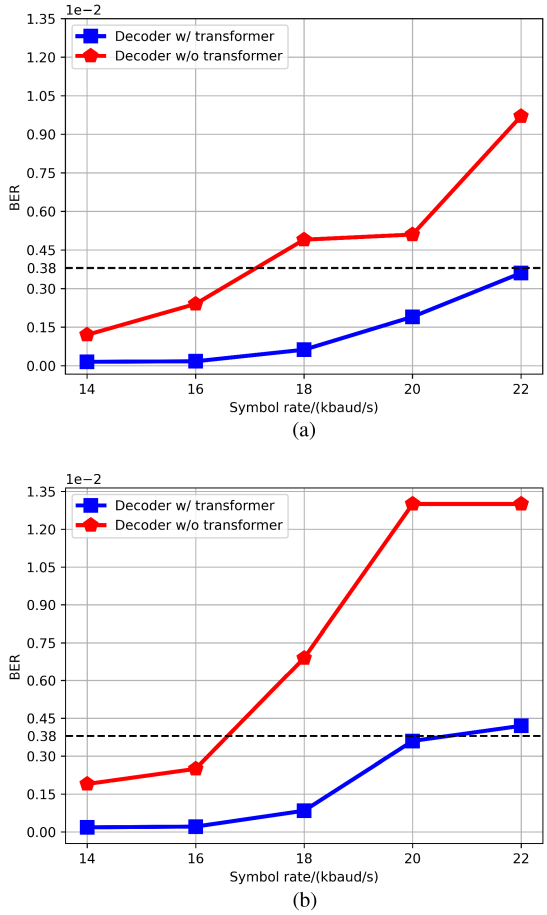


Fig. 11. (a) BER performance in small wave environment. (b) BER performance in big wave environment.

in the data rate. This is because the amplitude changes in the header are significant, which means that the negative impact of interference on decoding the header is not substantial. As a result, regardless of whether the transformer architecture is used or not, the model's ability to recognize the header remains similar, resulting in similar packet loss rates and data rates. However, the powerful capability of the transformer architecture to extract global features helps alleviate the negative effects of inter-symbol interference, leading to a significant decrease in the bit error rate. The experimental results demonstrate that the network using the transformer achieves more stable performance, leading to higher data rates and lower bit error rates.The experimental results showed that the performance of the network using the transformer was more stable, resulting in higher data rates and lower BER. Traditional convolutional neural networks (CNNs) use a fixed size local receptive field when processing images, potentially losing information from remote image locations. In contrast, the transformer's self-attention mechanism performs adaptive weight calculation for any position in the image, regardless of receptive field size. The weight matrix of the self-attention mechanism can be interpreted as the model's attention to different inputs and their relationship with one another. This enables better capture of the relationship between symbol stripes and their combined information in the time domain. Moreover, it helps alleviate the negative impact of inter-symbol

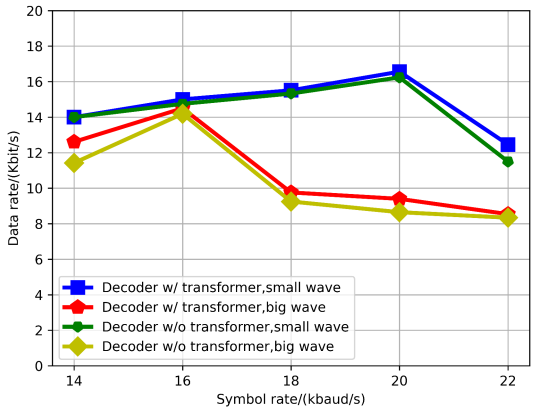


Fig. 12. Data rate comparison.

interference on decoding. By learning the sequence information of symbol stripe combinations, the network can improve the detection rate of data packet header. Better recognition of data packet headers can reduce the packet loss rate and improve the overall data rate. The demodulator without transformer is unable to recognize as many packet headers, resulting in more packet loss and a decrease in data rate.

To explore the performance of the token embedding based on SE-Res block, we conducted three sets of experiments in

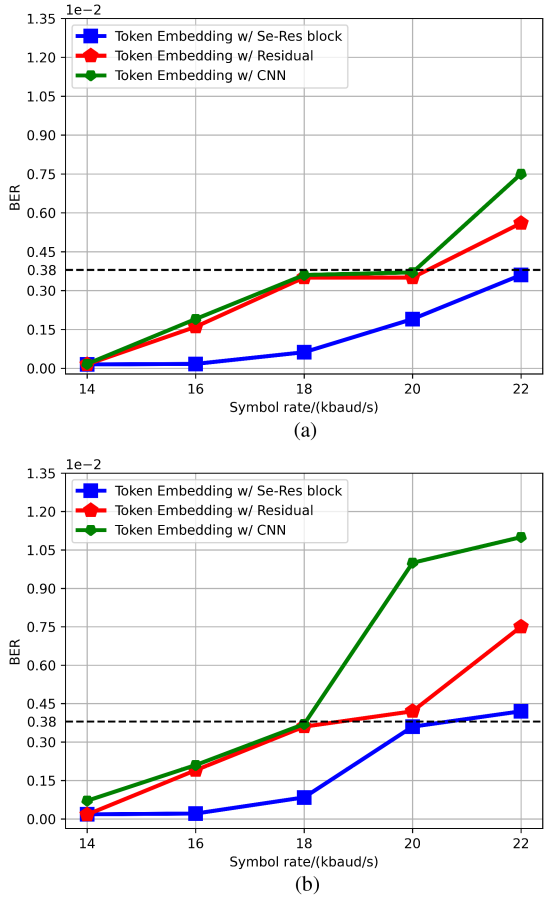


Fig. 13. (a) BER performance by using different Token-Embedding in small wave environment. (b) BER performance by using different Token-Embedding in big wave environment.

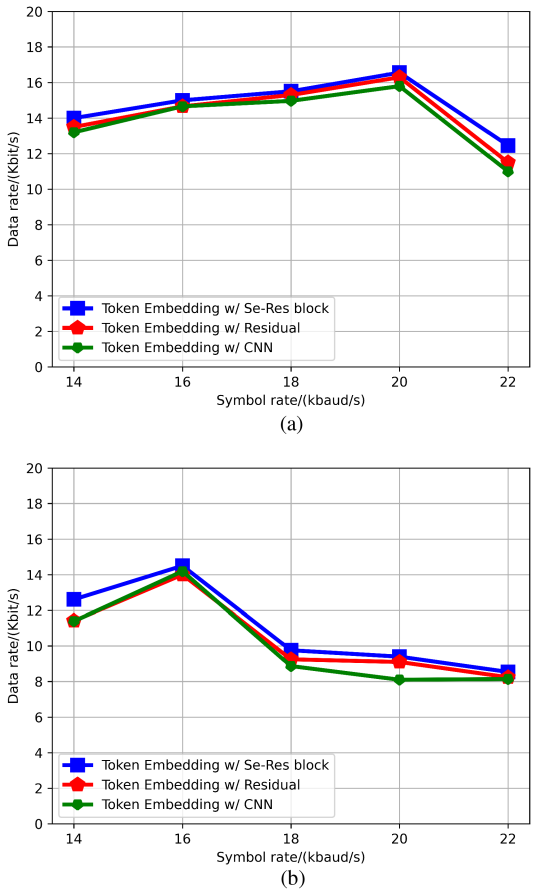


Fig. 14. (a) Data rate comparison by using different Token-Embedding in small wave environment. (b) Data rate comparison by using different Token-Embedding in big wave environment.

small wave and large wave environments, respectively. The first set used our proposed complete demodulation network model, the second set used residual operation for token embedding, and the third group used traditional CNN networks for token embedding with convolution layers, BN layers, ReLU activation function and pooling layers. Figs. 13 and 14 show the data rate and BER performance at symbol rates ranging from 14 kbaud to 22 kbaud. The results demonstrated that our proposed demodulation network model with token embedding based on SE-Res blocks outperformed the other two methods in terms of data rate and BER performance. The system using our demodulator exhibits a higher header recognition rate, which results in a faster data rate. In contrast, other architecture have weaker recognition capabilities for packet headers, thereby increasing the packet loss rate and decreasing the data rate. Moreover, token embedding using SE-Res blocks can more effectively extract features from the sub-images, thereby reducing the workload of subsequent network structures and improving the accuracy of classification results.

V. Conclusion

In this article, we proposed and demonstrated a water-air RSE-based optical camera system. We proposed a bubble-wave-mitigation algorithm to mitigate the negative impact of waves and bubbles. Moreover, to address the issue of performance deterioration caused by ISI, we transformed the decoding problem into a sequence classification problem and proposed a transformer-based neural network demodulator. The result showed that a data rate of 16.56 Kbit/s can be achieved under the FEC limit of $3.8 * 10^{-3}$ by applying the proposed scheme. Additionally, the results demonstrated that our approach improved the transmission performance of the water-air RGB-LED-based OCC system, serving as a valuable reference for future researchers.

References

[1] L.-K. Chen, Y. Shao, and Y. Di, "Underwater and water-air optical wireless communication," *J. Lightw. Technol.*, vol. 40, no. 5, pp. 1440–1452, Mar. 2022.

[2] X. Sun, M. Kong, C. Shen, C. H. Kang, T. K. Ng, and B. S. Ooi, "On the realization of across wavy water-air-interface diffuse-line-of-sight communication based on an ultraviolet emitter," *Opt. Exp.*, vol. 27, no. 14, pp. 19635–19649, 2019.

[3] S. Karp, "Optical communications between underwater and above surface (satellite) terminals," *IEEE Trans. Commun.*, vol. 24, no. 1, pp. 66–81, Jan. 1976.

[4] M. S. Islam and M. F. Younis, "Analyzing visible light communication through air–water interface," *IEEE Access*, vol. 7, pp. 123830–123845, 2019.

[5] J. Yen, J. Wang, S. Supittayapornpong, M. A. M. Vieira, R. Govindan, and B. Raghavan, "Meeting SLOs in cross-platform NFV," in *Proc. 16th Int. Conf. Emerg. Netw. Experiments Technol.*, 2020, pp. 509–523.

[6] J. Wang, T. Lévai, Z. Li, M. A. M. Vieira, R. Govindan, and B. Raghavan, "Quadrant: A cloud-deployable NF virtualization platform," in *Proc. 13th Symp. Cloud Comput.*, 2022, pp. 493–509.

[7] B. Islam, Y. Luo, and S. Nirjon, "Amalgamated intermittent computing systems," in *Proc. ACM/IEEE 8th Conf. Internet Things Des. Implementation. ACM*, 2023, pp. 184–196.

[8] Y. Luo, L. Zhang, Z. Wang, and S. Nirjon, "Efficient multitask learning on resource-constrained systems," 2023, *arXiv:2302.13155*.

[9] C. Yu, T. Yoo, H. Kim, T. T.-H. Kim, K. C. T. Chuan, and B. Kim, "A logic-compatible eDRAM compute-in-memory with embedded ADCs for processing neural networks," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 68, no. 2, pp. 667–679, Feb. 2021.

[10] Z. Zeng, "A survey of underwater wireless optical communication," Ph.D. dissertation, The Univ. British Columbia, 2015.

[11] Z. Zhou, S. Wen, Y. Li, W. Xu, Z. Chen, and W. Guan, "Performance enhancement scheme for RSE-based underwater optical camera communication using de-bubble algorithm and binary fringe correction," *Electronics*, vol. 10, no. 8, 2021, Art. no. 950.

[12] J. He and B. Zhou, "A deep learning-assisted visible light positioning scheme for vehicles with image sensor," *IEEE Photon. J.*, vol. 14, no. 4, pp. 1–7, Aug. 2022.

[13] M. Akram, L. Aravinda, M. Munaweera, G. Godaliyadda, and M. Ekanayake, "Camera based visible light communication system for underwater applications," in *Proc. IEEE Int. Conf. Ind. Inf. Syst.*, 2017, pp. 1–6.

[14] M. Akram, R. Godaliyadda, and P. Ekanayake, "Design and analysis of an optical camera communication system for underwater applications," *IET Optoelectron.*, vol. 14, no. 1, pp. 10–21, 2020.

[15] P. Luo, M. Zhang, Z. Ghassemlooy, S. Zvanovec, S. Feng, and P. Zhang, "Undersampled-based modulation schemes for optical camera communications," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 204–212, Feb. 2018.

[16] X. Sun et al., "A review on practical considerations and solutions in underwater wireless optical communication," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 421–431, Jan. 2020.

[17] C.-W. Chow, C.-Y. Chen, and S.-H. Chen, "Enhancement of signal performance in LED visible light communications using mobile phone camera," *IEEE Photon. J.*, vol. 7, no. 5, pp. 1–7, Oct. 2015.

[18] T. Nguyen, C. H. Hong, N. T. Le, and Y. M. Jang, "High-speed asynchronous optical camera communication using led and rolling shutter camera," in *Proc. IEEE 7th Int. Conf. Ubiquitous Future Netw.*, 2015, pp. 214–219.

[19] B. W. Kim, J.-H. Yoo, and S.-Y. Jung, "Design of streaming data transmission using rolling shutter camera-based optical camera communications," *Electronics*, vol. 9, no. 10, 2020, Art. no. 1561.

[20] R. L. P. de Lima, F. C. Boogaard, and R. E. de Graaf-van Dinther, "Innovative water quality and ecology monitoring using underwater unmanned vehicles: Field applications, challenges and feedback from water managers," *Water*, vol. 12, no. 4, 2020, Art. no. 1196.

[21] R. Deng, J. He, Y. Hong, J. Shi, and L. Chen, "2.38 Kbits/frame WDM transmission over a CVLC system with sampling reconstruction for SFO mitigation," *Opt. Exp.*, vol. 25, no. 24, pp. 30575–30581, 2017.

[22] O. I. Younus et al., "Data rate enhancement in optical camera communications using an artificial neural network equaliser," *IEEE Access*, vol. 8, pp. 42656–42665, 2020.

[23] W. Guan et al., "The detection and recognition of RGB-LED-ID based on visible light communication using convolutional neural network," *Appl. Sci.*, vol. 9, no. 7, 2019, Art. no. 1400.

[24] P. Ling, M. Li, and W. Guan, "Channel-attention-enhanced LSTM neural network decoder and equalizer for RSE-based optical camera communications," *Electronics*, vol. 11, no. 8, 2022, Art. no. 1272.

[25] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[26] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[27] F. Wen, M. Qin, P. Gratz, and N. Reddy, "OpenMem: Hardware/software cooperative management for mobile memory system," in *Proc. IEEE/ACM 58th Des. Automat. Conf.*, 2021, pp. 109–114.

[28] F. Wen, M. Qin, P. V. Gratz, and A. L. N. Reddy, "Hardware memory management for future mobile hybrid memory systems," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 39, no. 11, pp. 3627–3637, Nov. 2020.

[29] Y. Liu et al., "Comparison of thresholding schemes for visible light communication using mobile-phone image sensor," *Opt. Exp.*, vol. 24, no. 3, pp. 1973–1978, 2016.

[30] F. Zhou, Z. Fu, and D. Zhang, "High dynamic range imaging with context-aware transformer," *Int. Joint Conf. Neural Netw.*, 2023, pp. 1–8.

[31] D. Zhang and F. Zhou, "Self-supervised image denoising for real-world images with context-aware transformer," *IEEE Access*, vol. 11, pp. 14340–14349, 2023.

[32] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 213–229.

[33] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[35] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 2488–2498.

[36] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertainty, Fuzziness Knowl. Based Syst.*, vol. 6, no. 2, pp. 107–116, 1998.

[37] S. Hochreiter et al., "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," in *A Field Guide to Dynamical Recurrent Neural Networks*. Piscataway, NJ, USA: IEEE Press, 2001.

[38] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[39] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.