Research Article Vol. 32, No. 2 / 15 Jan 2024 / Optics Express 1489
Optics EXPRESS

Underwater optical signal detection system using diffuser-based lensless imaging

YINUO HUANG, GOKUL KRISHNAN, SAURABH GOSWAMI, AND BAHRAM JAVIDI*

Electrical and Computer Engineering Department, University of Connecticut, 371 Fairfield Way Unit 4157, Storrs, Connecticut 06269, USA
bahram.javidi@uconn.edu

Abstract: We propose a diffuser-based lensless underwater optical signal detection system. The system consists of a lensless one-dimensional (1D) camera array equipped with random phase modulators for signal acquisition and one-dimensional integral imaging convolutional neural network (1DInImCNN) for signal classification. During the acquisition process, the encoded signal transmitted by a light-emitting diode passes through a turbid medium as well as partial occlusion. The 1D diffuser-based lensless camera array is used to capture the transmitted information. The captured pseudorandom patterns are then classified through the 1DInImCNN to output the desired signal. We compared our proposed underwater lensless optical signal detection system with an equivalent lens-based underwater optical signal detection system in terms of detection performance and computational cost. The results show that the former outperforms the latter. Moreover, we use dimensionality reduction on the lensless pattern and study their theoretical computational costs and detection performance. The results show that the detection performance of lensless systems does not suffer appreciably. This makes lensless systems a great candidate for low-cost compressive underwater optical imaging and signal detection.

1. Introduction

Underwater optical signal detection systems (UOSDS) are receiving growing interest among researchers [1–3] for their crucial role in applications such as ocean exploration, defense, and environmental monitoring. However, owing to the intrinsic and extrinsic properties of underwater environments, such as scattering, absorption, and turbulence [4], building a low bit-error-rate and high-speed optical signal detection system is a challenging task. Moreover, possible underwater occlusion can lead to a higher bit-error rate for UOSDS. Recently, to overcome these issues, many approaches, such as multi-modal sensing [5], diversity-reception-based optical communication [6], and three-dimensional-integral-imaging-based (3D InIm) optical signal detection systems [7–9], have been proposed. Among these approaches, 3D InIm-based approaches [7,9] could achieve excellent detection performance under degradations such as partial occlusion and turbidity.

Traditionally, imaging devices have mostly consisted of optical elements such as lenses, mirrors, etc., and image sensors. Lensless imaging is a relatively newer idea that replaces lenses with a phase modulator [10], a programmable modulator [11], or an amplitude mask modulator [12]. Lensless imaging, compared to conventional lens-based systems, carries several advantages, such as compact size, smaller weights, and larger field of view (FoV) [13]. In addition, since lensless systems scatter the incoming information widely over the sensor as opposed to a lens that focuses the information on small areas on the sensor, it offers better data compressibility for applications such as storage and classification [14]. As such, lensless imaging-based approaches started becoming popular in many applications, such as microscopy [15], 3D sensing [16], and photography [12]. Moreover, with the advent of deep learning-based approaches, we could realize cell classification [17] and disease classification [14] without complicated reconstruction of images with better accuracy and at a lower cost.

#512438 https://doi.org/10.1364/OE.512438
Journal © 2024 Received 15 Nov 2023; accepted 18 Dec 2023; published 2 Jan 2024

Optics EXPRESS

Even though 3D InIm-based approaches [7–9] perform well in degraded environments, they incur a high computational cost. To address this, the one-dimensional integral imaging convolutional neural network (1DInImCNN) [18] has been proposed, which, without performing 3D InIm reconstruction, makes use of information from multiple perspectives and classifies optical signals through an end-to-end pipeline. Thus, 1DInImCNN outperforms the conventional 3D InIm-based approach [7] in both detection performance and computational cost. In the current work, inspired by [14], we propose a novel diffuser-based lensless underwater optical signal detection system (UOSDS) that combines lensless 1D camera array with 1DInImCNN to minimize the computational cost and achieve better detection performance as compared to the lens-based UOSDS. For our experiments, a lensless 1D camera array is used to capture the temporally encoded optical signal transmitted using a light-emitting diode (LED). A water tank filled with turbid water and artificial plants is used to mimic the degraded underwater environment. The 1DInImCNN processes the videos captured by the lensless imaging system and outputs the signal symbols. Our results show that the lensless UOSDS outperforms the lens-based UOSDS in terms of detection performance. Furthermore, dimensionality reduction techniques are applied to the lensless UOSDS, and the results show that our dimensionality-reduced lensless system is capable of achieving considerably less computational cost compared to the lens-based systems without affecting performance much.

The rest of the paper is organized as follows: Section 2 covers a brief review of the 1DInImCNN approach, experimental setup and data collection, and procedures for applying dimensionality reduction. The results and discussions regarding theoretical computational costs and detection performance are included in Section 3. Finally, Section 4 concludes this paper.

2. Methodology

The proposed lensless UOSDS aims to achieve performance enhancement and cost efficiency compared to the traditional lens-based UOSDS. Figure 1 shows the block diagram for the proposed system. Here, we used an end-to-end optical signal detection system that combines image acquisition and signal detection without the need for an intermediate reconstruction stage. For image acquisition, diffuser-based lensless cameras aligned in the 1D configuration are used. For signal detection, 1DInImCNN [18] is used to classify the videos captured by the lensless 1D camera array. The inputs to 1DInImCNN are the videos that contain pseudorandom patterns caused by the illumination from the optical transmitter. The outputs of the 1DInImCNN are the decoded signals.

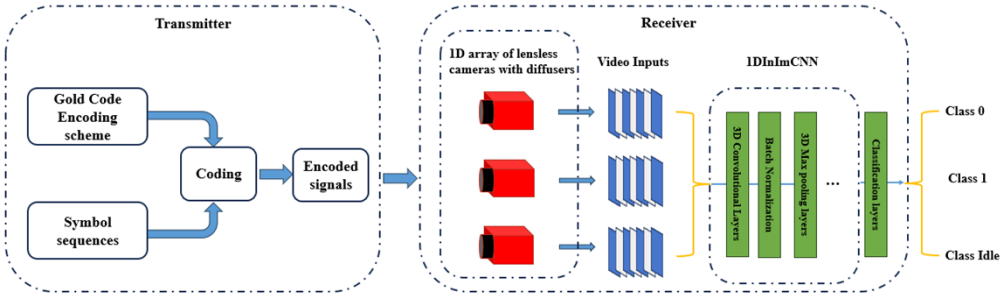

Fig. 1. Block diagram for the proposed lensless underwater optical signal detection system (UOSDS).

Optics EXPRESS

2.1. Experimental setup

Figure 2 shows the experimental setup. We used a 630 nm light-emitting diode (LED) for transmitting the optical signals. A 7-bit gold code sequence encodes the symbol sequences into signal sequences. Specifically, in the experiment, the encoded signal sequence [1,1, 0, 0, 1, 0, 1] represents the symbol "1" and the flipped sequence [0, 0, 1,1, 0, 1, 0] represents the symbol "0" during the transmission. The transmitter is programmed using an Arduino board to send the encoded signal sequences at the frequency of 20 Hz. We use the "LED on" condition to represent signal bit "1" and "LED off" to represent signal bit "0".

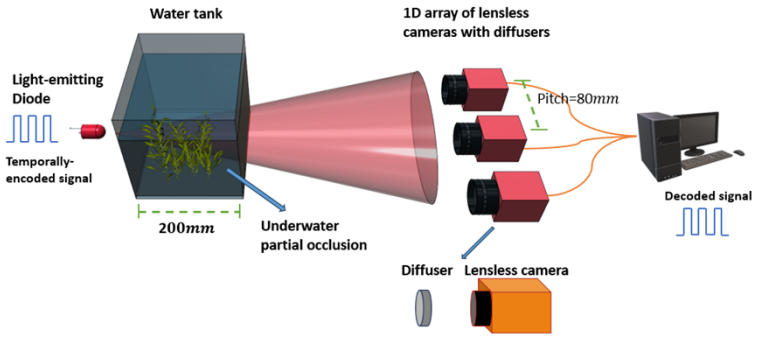

Fig. 2. Experimental setup for the lensless underwater optical signal detection system (UOSDS). Three lensless cameras are aligned and each equipped with a diffuser. An underwater environment with turbidity and partial occlusion is created inside a water tank.

A water tank of dimension $200mm \times 200mm \times 200mm$ is placed in front of the transmitter to create a degraded underwater environment. Degraded environments such as turbid water and partial occlusion are created by adding anti-acid and placing artificial plants into the water tank. The Beer-Lambert's law, $I = I_o e^{-\alpha z}$, is used to measure the turbidity as done in [8]. Here, I is the intensity of the light after traveling z distance in the turbid medium. I_o is the initial intensity before traveling in the turbid medium. z is the distance, and α is the attenuation coefficient in units of mm^{-1}.

A one-dimensional lensless camera array equipped with holographic diffusers constitutes the receivers for the transmitted optical signals. Three Mako G-319 cameras are aligned in a one-dimensional configuration, shown in Fig. 2. The pitch between each camera is 80 mm. Holographic diffusers are used to replace the traditional zoom lens. The diffusers have a thickness of 0.78 mm and are applicable in the wavelength range of 400 to 700 nm. The diffusing angle is 0.5 degrees, measured in full width half maximum (FWHM), and the substrate is made of polycarbonate. The holographic diffusers are cut in a circular shape with a diameter of 17 mm in order to be placed in front of the camera. During data collection, three cameras are synchronized, and the frame rate is set to the transmitted signal frequency to achieve synchronization [19] between the transmitter and receivers. The bit rate of the proposed optical signal detection system can be potentially improved through the advancement in high-speed cameras [20], which can achieve more than 1,000,000 frames per second. The nature of the lens-based imaging systems is that lenses can converge light, as opposed to the nature of the lensless diffuser-based imaging systems that scatter light across the sensors. Intuitively, the maximum pixel intensity for lensless imaging systems is much lower than that for lens-based imaging systems. Therefore, to have equivalent lensless imaging systems compared to lens-based imaging systems, we maintain the same maximum pixel intensity for both systems in the experiment. The camera's resolutions are set to 1600×1200 pixels, and collected videos are resized to 400×300 pixels to ease the processing burden for the neural network during the training and testing stage.

Optics EXPRESS

2.2. One-dimensional integral-imaging convolutional neural network (1DInImCNN)

In this paper, we use 1DInImCNN [18] due to its superior performance in underwater optical signal detection applications. Figure 3 shows the structure for 1DInImCNN. It consists of multiple input layers, each with size [x, y, d, c]. Here, x and y represent the lateral dimensions of the input videos. d is the number of frames for each input video, and c is the number of color channels for input videos. d is set to 7 since we use a 7-bit encoding scheme, and c is set to 3 because we used a standard RGB camera to capture the videos. A depth concatenation layer that combines the input videos in the third dimension (d-dimension) integrates different input videos into one matrix for feature extraction and classification. A 3D convolution layer containing 32 kernels of size [7,7,7] and a stride of [5,5,7] is used to extract information from each of the input videos. The 3D convolution layer is followed by a 3D max-pooling layer with a pool size of [4,4,4] and stride size of [2,2,2] to downsample the feature maps. A proper max-pooling layer helps to select the sharpest features, resulting in possible increasing classification performance, and can reduce the size of the feature, hence decreasing the computational cost. After the 3D max-pooling layer, three 3D convolutional layers that have 64, 128, and 64 kernels, respectively, follow to extract the high-level features. They have kernel sizes of [5,5,5], [3,3,2], and [1,1,1], respectively, with a stride of [3,3,1], [2,2,1] and [1,1,1] respectively. Widely accepted modules such as batch normalization [21] and ReLU activation functions follow each 3D convolution layer in 1DInImCNN. Finally, a fully connected layer with an output size of 3 follows the last 3D convolution layer to output the possible three classes from the videos. For three-input 1DInImCNN, the input videos from 3 cameras are utilized. For comparison with 2D imaging (single camera), we use the same network architecture but only with a single input. Videos captured by the center camera are used for 2D imaging.

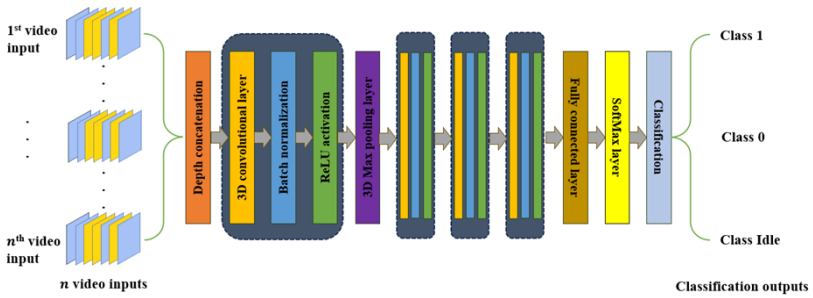

Fig. 3. Structure for n-input 1DInImCNN

2.3. Data collection and generation of training and testing data

Figure 4(a) shows the experimental condition during the collection of the training data. We collected data under five gradually increasing levels of turbidities. For a particular level of turbidity, we transmit four 8-symbol sequences [1, 0, 0, 1, 1, 0, 1, 0], each of which is encoded using the 7-bit Gold code sequence mentioned above. From captured videos, we can make training videos for three classes: class 0, class 1, and class idle (i.e., data belonging to neither of the other two classes). For training videos in class 0, each video has a number of frames equal to the length of the encoding scheme, and the frames capture the bit information [0, 0, 1, 1, 0, 1, 0] in sequential order. Similarly, for training videos in class 1, the frames should capture the bit information [1, 1, 0, 0, 1, 0, 1] in sequential order. For videos in the class idle, they should include all possibilities that might arise from all possible 7-bit sequences. Due to the particular encoding scheme used in this work, there are only 36 possible signals that might be generated. Excluding the signal for class “0” and class “1”, we are left with only 34 combinations. We make

Optics EXPRESS

one video for each combination among 34 combinations, and we have 34 videos in class idle. To deal with the imbalanced training dataset, we apply random minority oversampling to class 1 and class 0 to increase the size of the training data. Thus, for one particular turbidity, we have a total of 102 videos. We repeat the procedure for five turbidity levels. Table 1 shows the attenuation coefficients for training data. Figure 4(b) and Fig. 4(c) show sample images that capture the LED "on" state in $\alpha = 0.0037\ mm^{-1}$ and $0.0095\ mm^{-1}$, respectively.

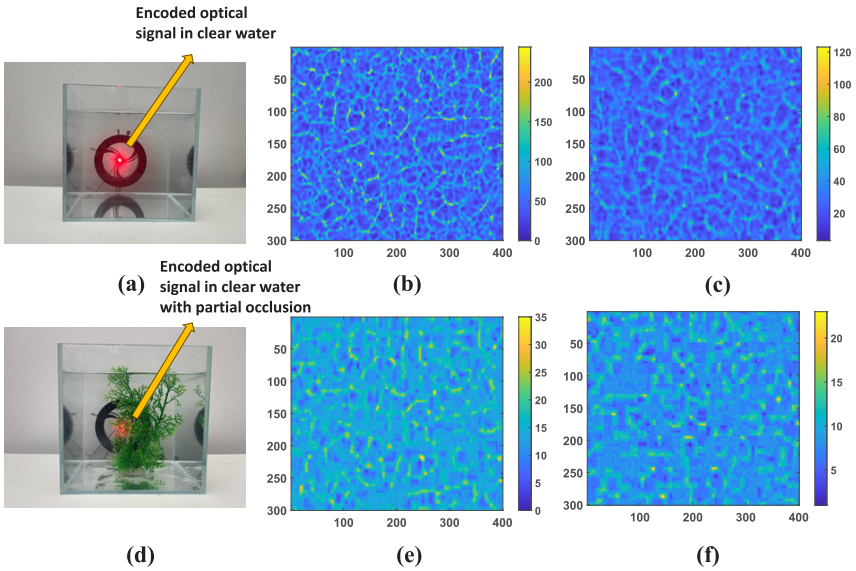

Fig. 4. Experimental setup for training and testing data collection and sample collected images. (a): experimental condition for collection of training videos. (b-c): sample images of training data captured by the center camera from a lensless 1D camera array at α=0.0037 mm^{-1} and 0.0095 mm^{-1}, respectively. (d): experimental condition for collection of testing videos. (e-f): sample images of testing data captured by the center camera from a lensless 1D camera array at α=0.0170 mm^{-1} and 0.0198 mm^{-1}, respectively.

Table 1. Turbidity levels (α in mm^{-1}) for training data

	Clear water	Turbid Level 1	Turbid Level 2	Turbid Level 3	Turbid Level 4
Attenuation coefficient $\boldsymbol{\alpha}$	0.0007	0.0037	0.0064	0.0095	0.0145

Figure 4(d) shows the experimental setup for collecting the testing data. Regarding testing data collection, we placed artificial plants inside the water tank to mimic the occlusion in the underwater environment. We have adjusted eight levels of turbidity to test the performance of the trained classifiers, and Table 2 shows the attenuation coefficients used for testing data. Figure 4(e) and Fig. 4(f) show sample images from testing data that capture the LED "on" in $\alpha = 0.0170$ mm^{-1} and $0.0198\ mm^{-1}$, respectively. During the collection of testing data, under a particular level of turbidity, we use the LED to transmit 64 symbols of sequence and use the lensless camera array to capture this 448-bit long signal sequence. To make testing videos for each class, we use a sliding window approach to sequentially slice the long video into 7-frame videos. Figure 5 shows the process of applying the sliding window approach. Suppose we have a long video with k frames, and we label each frame sequentially from 1 to k. To make the first video, we used frame number 1 to frame number 7. After the first video, we used the images from frame number 2 to frame number 8 to make the second video. By repeating this procedure, the last video is

Optics EXPRESS

made from frame number $k-6$ to frame number k. After slicing, we have 32 videos for both class 0 and class 1, and we have 272 videos in class idle.

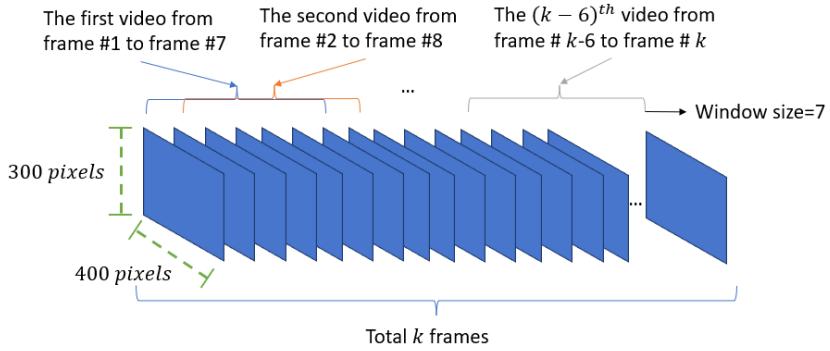

Fig. 5. Demonstration of sliding window approach for test data.

Table 2. Turbidity levels (α in mm^{-1}) for testing data. α is the attenuation coefficient.

	Attenuation coefficient α in mm^{-1}
Clear water	0.0010
Turbid Level 1	0.0032
Turbid Level 2	0.0071
Turbid Level 3	0.0170
Turbid Level 4	0.0176
Turbid Level 5	0.0198
Turbid Level 6	0.0215

2.4. Dimensionality reduction

In [14], researchers show that, by applying dimensionality reduction to the images taken from a lensless single random phase encoding (SRPE) system, the resultant one-dimensional (1D) or two-dimensional (2D) strips of pixels can still retain information adequate for carrying out classification. In our experiment, we also include the performance of the classifiers trained using the dimensionality-reduced training data. The optimally trained classifiers are evaluated using the dimensionality-reduced testing data generated from the original testing data. Figure 6 shows example images before and after applying dimensionality reduction. Dimensionality reduction is applied to videos by randomly taking the 1D rows or columns of pixels. To be specific, the dimensionality-reduced training videos have the dimension of $[1,\ v,\ d,\ c]$ or $[v,\ 1,\ d,\ c]$ when we want to take a 1D vector of size v from the original training data. For our experiment, we used three values for v (50, 100, and 150 pixels). To generate the dimensionality-reduced training videos, we will randomly pick one location from the original training videos, assuming $[j,\ k,\ 1,\ 1]$. Then, we will randomly choose the horizontal or vertical orientation to apply a line crop sized v pixels. After determining the location and orientation of the 1D vector, we can generate the training data using $[j:(j+v-1),\ k,\ 1:d,\ 1:c]$ or $[j,\ k:(k+v-1),\ 1:d,\ 1:c]$ matrix operations. For training videos from 3-input 1DInImCNN, the three input videos need to use the same location and orientation to generate three dimensionality-reduced training videos. We repeat the same procedure for every instance in the original training data to generate the dimensionality-reduced training data. The 1D adaptation of the same neural network architecture

Optics EXPRESS

in Fig. 3 has been utilized for training, hyperparameter tuning and testing the network. Specifically, for 1D adaptation of the 3-input 1DInImCNN, we set the first dimension of kernel sizes and stride sizes to 1 for all the layers. We repeat the same procedure for other networks.

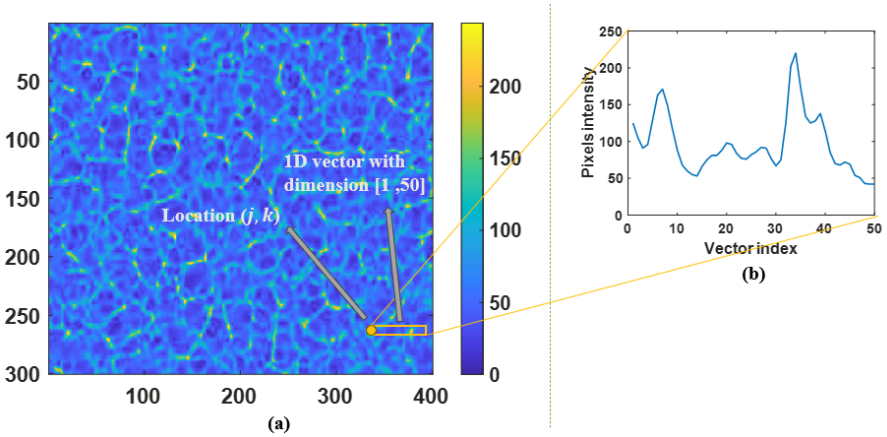

Fig. 6. (a) Sample image from the training data before applying dimensionality reduction. (b) The cropped vector after applying the dimensionality reduction to the corresponding sample image using a 1D vector with a size of 50 pixels. The y-axis of (b) shows the pixel's intensity and the x-axis shows the vector index.

To generate the dimensionality-reduced testing data, we use the same procedure as done with the training dataset. However, for the test dataset, from a single testing video, we picked 20 random combinations of positions and orientations to generate 20 dimensionality-reduced testing videos. This has been done to get an averaged performance by eliminating small discrepancies in performance across different locations and orientations. The size of dimensionality-reduced testing data for one particular turbidity for class 0, class 1, and class idle are 640, 640, and 5440, respectively.

3. Result and discussion

We compare the performance of the proposed lensless system with an equivalent lens-based system and a dimensionality-reduced lensless system. Theoretical analysis of the computational cost and detection performance analysis has been done. The detection performance and computational cost for lens-based underwater optical signal detection with dimensionality reduction are not considered in our experiments because the lens-based systems, unlike lensless systems, do not scatter the information of the transmitted signal across the entire sensors.

3.1. Analysis of detection performance and computational cost

In this subsection, we compare the detection performance and computational cost between 2 different systems: the lensless UOSDS and the lens-based UOSDS. 1DInImCNN is used in both systems to classify the optical signals. For lens-based UOSDS, we used the procedure from the previously published paper [18] to collect the training and testing data. For the networks from both systems, we use Bayesian optimization to select the optimal hyperparameters. Training data and validation data are split into 90% and 10%, and classification accuracy is used as the metric to select the optimal classifiers. Mini-batch size is optimized in the range between 4 to 50. The initial learning rate is optimized in the range between 0.001 and 0.001 in the log scale. L2 regularization is optimized in the range between 10^{-10} to 0.01 in the log scale. We trained all the

Optics EXPRESS

networks in one computer with the following configuration: Intel i9-10940X CPU and Nvidia Quadro RTX A5000.

To evaluate the classification performance of a multiclass classification problem with the imbalanced dataset, we choose to use the Matthew correlation coefficient (MCC) as the main metric. Compared to metrics such as area under the curve (AUC) and receiver operating characteristic (ROC), MCC can reveal more information regarding the classification performance [22,23]. Also, MCC is widely adopted in evaluating classification performance in machine learning literature [24–26]. The values of MCC should be greater or equal to -1 and less than or equal to 1. 1, 0, and -1 represent the perfect classification, random classification, and worst classification, respectively.

Figure 7 shows the classification performance of the lensless and lens-based systems across different turbidity levels. The number of pixels for each camera are 400×300 pixels or 1×50 pixels as indicated in Fig. 7. For the dimensionality-reduced diffuser-based multi-camera configuration with 1×50 pixels per camera, we compute the MCC value for a particular turbidity by averaging MCC values across 20 different testing videos using dimensionality reduction from one particular full-resolution test video. From Fig. 7, all approaches result in perfect detection for $\alpha \leq 0.0176\ mm^{-1}$. However, as turbidity increases, the detection performance for all approaches start decreasing. At higher turbidities ($\alpha = 0.0198\ mm^{-1}$ and $0.0215\ mm^{-1}$), the classifier trained with the lensless dataset can outperform the classifier trained with the lens-based dataset for both multi-camera configuration and single-camera configuration. Moreover, the multi-camera system has better performance compared to the single camera system for both lensless and lens-based system. The classification performance decreases with decrease in the number of pixels. However, the dimensionality-reduced multi-camera configuration still has classifiable information, and it reaches perfect detection performance for $\alpha \leq 0.0176\ mm^{-1}$ (see Fig. 7). Also, by comparing lens-based multi-camera configuration with 400×300 pixels per camera to dimensionality-reduced diffuser-based multi-camera configuration in Fig. 7, we can infer that the performance of the dimensionality-reduced classifier is even comparable to the classifier which uses lens-based full-resolution videos. In α=0.0215, the dimensionality-reduced classifier can slightly outperform the lens-based multi-camera configuration with 400×300 pixels.

The advantages for the classifiers trained using a dimensionality-reduced dataset achieve not only comparable performance but also a great reduction in the computational cost. To theoretically analyze the computational cost, the number of floating point operations (FLOPs) is used as a metric. FLOPs are widely used in machine learning literature to measure the complexity of neural networks [27–29]. In [18], the derived equations are shown for 3D convolutional layers, 3D max-pooling layers, ReLU layers, and fully connected layers, and they can be used to analyze the FLOPs used in our neural networks. The number of FLOPs used for lens-based multi-camera configuration with 400×300 pixels and dimensionality-reduced diffuser-based multi-camera configuration with 1×50 pixels are 1.114×10^9 and 4.649×10^6. The lowest dimensionality-reduced lensless UOSDS considered is 240 times better in computational costs compared to the lens-based UOSDS.

Figures 8 and 9 show the boxplots that summarizes the results for lensless diffuser-based system with dimensionality-reduced configurations. The boxplot is generated using 140 MCC values from seven different turbidity levels and 20 different dimensionality-reduced configurations for each turbidity level. Table 3 summarizes the statistics derived from Fig. 8 and Fig. 9. Figure 8, Fig. 9, and Table 3 shows that the detection performance is enhanced with increased number of pixels. Since three camera configuration utilizes thrice the number of pixels than a single camera configuration, we considered equal number of pixels for single camera and three camera configurations for fairness. Thus, in addition to comparing three camera 1×50 configuration with a single camera 1×50 configuration, we also consider single camera 1×150 configuration that has the same number of pixels as that of the three camera 1×50 configuration.

Optics EXPRESS

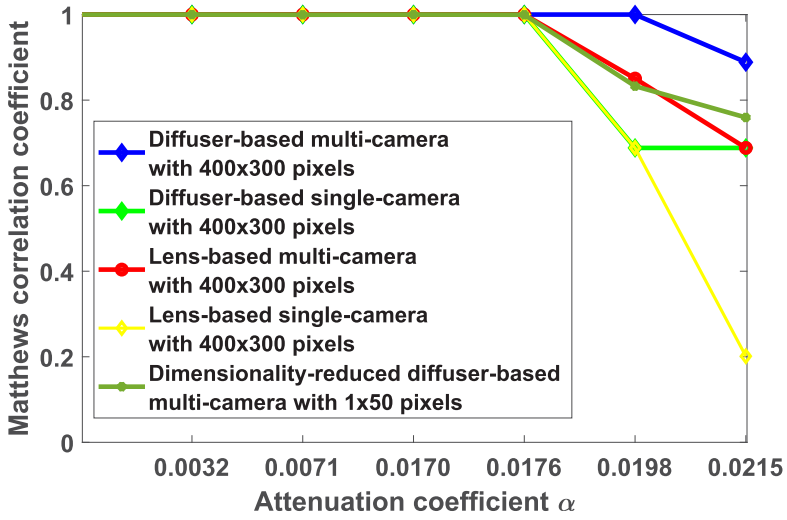

Fig. 7. Matthew correlation coefficient (MCC) for the proposed diffuser-based lensless underwater optical signal detection system (UOSDS), previously proposed lens-based UOSDS, dimensionality-reduced lensless underwater optical signal detection system (UOSDS), and dimensionality-reduced diffuser-based lensless system across different turbidity levels. The MCCs for dimensionality-reduced diffuser-based multi-camera with 1×50 pixels are calculated by averaging 20 MCCs since we have generated 20 different dimensionality-reduced test videos from one full resolution test video. Here, multi-camera uses three cameras.

The minimum, 25th percentile, and the variance in Fig. 9 and Table 3 show that the lensless single camera configuration with 1×150 pixels has worse detection performance compared to the lensless three camera configuration with 1×50 pixels. Furthermore, variance in the detection performance (see Table 3) lowers as we increase the number of pixels in lensless imaging across all the configurations considered thus far. Thus, for lensless system, we conclude that the detection performance increases and becomes more stable with increasing number of pixels. Also, multi-camera system outperforms equivalent (same number of pixels) single camera system in both detection performance and computational cost.

Table 3. Summary of computational costs (based on FLOPs) and statistics derived from Fig. 8 and Fig. 9 for the lensless diffuser-based system with dimensionality-reduced configurations.

	Minimum	Variance	25th percentile	Medium	Maximum	FLOPs
Single camera 1×50 pixels	0.351	0.0626	0.523	1	1	2.00×10^6
Single camera 1×150 pixels	0.684	0.0120	0.796	1	1	5.47×10^6
Three camera 1×50 pixels per camera	0.729	0.0091	0.850	1	1	4.65×10^6
Three camera 1×100 pixels per camera	0.841	0.0023	0.931	1	1	9.30×10^6

Optics EXPRESS

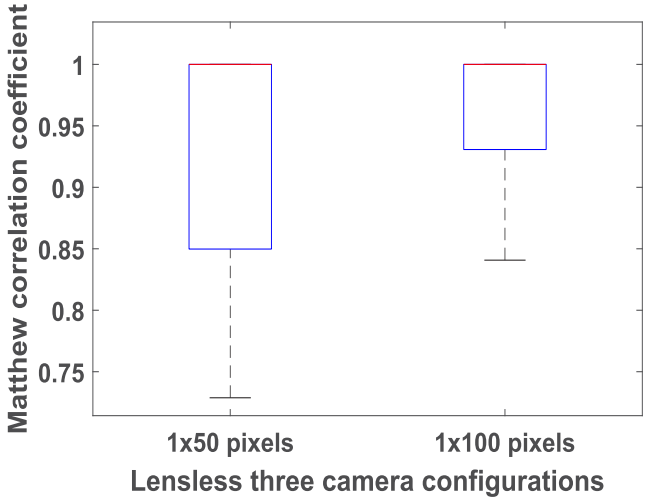

Fig. 8. Boxplots summarizing the performance using Matthew correlation coefficient for dimensionality-reduced classifiers in lensless three camera configurations. The comparison is performed between 1×50 pixels per camera and 1×100 pixels per camera. The lower bound of the blue box indicates the 25th percentile for all MCC values in the box. The black bar indicates the minimum value and the red bar indicates the medium value.

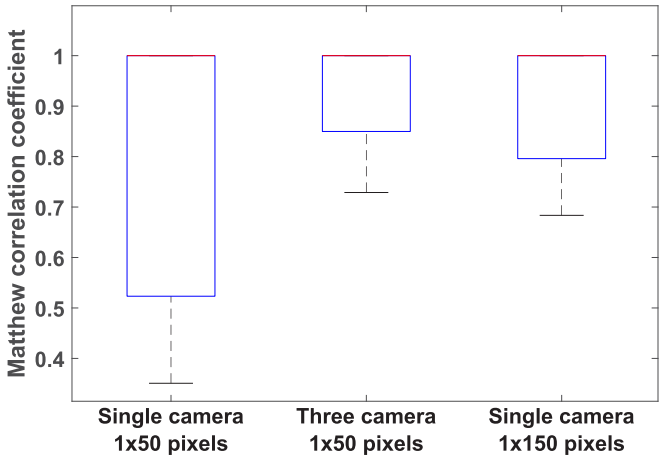

Fig. 9. Comparison of single camera and three camera lensless configurations with dimensionality reduction. Here, the comparison is performed between three camera configuration with 1×50 pixels per camera, single camera configuration with 1×50 pixels per camera, and single camera with 1×150 pixels per camera. The lower bound of the blue box indicates the 25th percentile for all MCC values in the box. The black bar indicates the minimum value and the red bar indicates the medium value.

4. Conclusion

In conclusion, we have proposed an underwater optical signal detection system that combines lensless imaging, 1D camera array, and 1DInImCNN. We compared the proposed system with previously proposed lens-based 1DInImCNN. The results show that our proposed lensless UOSDS can outperform the lens-based UOSDS under the experimental conditions considered. Moreover, we studied the dimensionality reduction in lensless underwater optical signal detection systems

Optics EXPRESS

and analyzed its performance. The results show that, by applying the dimension reduction to the lensless UOSDS, the computational cost can be significantly minimized without degrading the detection performance much.

Funding. Air Force Office of Scientific Research (FA9550-21-1-0333); Office of Naval Research (N000142212349, N000142212375).

Acknowledgments. We wish to acknowledge support under the Office of Naval Research (ONR) (ONR N000142212349, N000142212375); Air-Force Office of Scientific Research (AFOSR) (FA9550-21-1-0333).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time.

Reference

1. H.-H. Lu, C.-Y. Li, H.-H. Lin, *et al.*, "An 8 m/9.6 Gbps Underwater Wireless Optical Communication System," IEEE Photonics J **8**(5), 1–7 (2016).
2. H. Zhou, M. Zhang, X. Wang, *et al.*, "Design and Implementation of More Than 50 m Real-Time Underwater Wireless Optical Communication System," J. Lightwave Technol. **40**(12), 3654–3668 (2022).
3. D. Wen, W. Cai, and Y. Pan, "Design of underwater optical communication system," in *OCEANS 2016 - Shanghai* (2016), pp. 1–4.
4. X. Yi, Z. Li, and Z. Liu, "Underwater optical communication performance for laser beam propagation through weak oceanic turbulence," Appl. Opt. **54**(6), 1273–1278 (2015).
5. F. Campagnaro, F. Guerra, P. Casari, *et al.*, "Implementation of a multi-modal acoustic-optical underwater network protocol stack," in *OCEANS 2016 - Shanghai* (2016), pp. 1–6.
6. A. C. Boucouvalas, K. P. Peppas, K. Yiannopoulos, *et al.*, "Underwater Optical Wireless Communications With Optical Amplification and Spatial Diversity," IEEE Photonics Technol. Lett. **28**(22), 2613–2616 (2016).
7. G. Krishnan, R. Joshi, T. O'Connor, *et al.*, "Optical signal detection in turbid water using multidimensional integral imaging with deep learning," Opt. Express **29**(22), 35691–35701 (2021).
8. R. Joshi, G. Krishnan, T. O'Connor, *et al.*, "Signal detection in turbid water using temporally encoded polarimetric integral imaging," Opt. Express **28**(24), 36033 (2020).
9. R. Joshi, T. O'Connor, X. Shen, *et al.*, "Optical 4D signal detection in turbid water by multi-dimensional integral imaging using spatially distributed and temporally encoded multiple light sources," Opt. Express **28**(7), 10477–10490 (2020).
10. V. Boominathan, J. K. Adams, J. T. Robinson, *et al.*, "PhlatCam: Designed Phase-Mask Based Thin Lensless Camera," IEEE Trans. Pattern Anal. Mach. Intell. **42**(7), 1618–1629 (2020).
11. G. Huang, H. Jiang, K. Matthews, *et al.*, "Lensless Imaging by Compressive Sensing," in *2013 IEEE International Conference on Image Processing, ICIP 2013 - Proceedings* (2013).
12. M. S. Asif, A. Ayremlou, A. Sankaranarayanan, *et al.*, "FlatCam: Thin, Lensless Cameras Using Coded Aperture and Computation," IEEE Trans. Comput. Imaging **3**(3), 384–397 (2017).
13. V. Boominathan, J. T. Robinson, L. Waller, *et al.*, "Recent advances in lensless imaging," Optica **9**(1), 1–16 (2022).
14. P. M. Douglass, T. O'Connor, and B. Javidi, "Automated sickle cell disease identification in human red blood cells using a lensless single random phase encoding biosensor and convolutional neural networks," Opt. Express **30**(20), 35965–35977 (2022).
15. O. Mudanyali, D. Tseng, C. Oh, *et al.*, "Compact, Light-weight and Cost-effective Microscope based on Lensless Incoherent Holography for Telemedicine Applications," Lab Chip **10**(11), 1417–1428 (2010).
16. N. Antipa, G. Kuo, R. Heckel, *et al.*, "DiffuserCam: lensless single-exposure 3D imaging," Optica **5**(1), 1–9 (2018).
17. T. O'Connor, C. Hawxhurst, L. M. Shor, *et al.*, "Red blood cell classification in lensless single random phase encoding using convolutional neural networks," Opt. Express **28**(22), 33504–33515 (2020).
18. Y. Huang, G. Krishnan, T. O'Connor, *et al.*, "End-to-end integrated pipeline for underwater optical signal detection using 1D integral imaging capture with a convolutional neural network," Opt. Express **31**(2), 1367–1385 (2023).
19. W. Liu and Z. Xu, "Some practical constraints and solutions for optical camera communication," Phil. Trans. R. Soc. A. **378**(2169), 20190191 (2020).
20. P. Xia, Y. Awatsuji, K. Nishio, *et al.*, "One million fps digital holography," Electron Lett **50**(23), 1693–1695 (2014).
21. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach, eds., Proceedings of Machine Learning Research (PMLR, 2015), 37, pp. 448–456.
22. D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics **21**(1), 6 (2020).
23. G. Jurman, S. Riccadonna, and C. Furlanello, "A Comparison of MCC and CEN Error Measures in Multiclass Prediction," PLoS One **7**(8), e41882 (2012).
24. J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *2017 International Conference on Computing Networking and Informatics (ICCNI)* (2017), pp. 1–9.

Optics EXPRESS

25. Z. Liang, A. Powell, I. Ersoy, *et al.*, "CNN-based image analysis for malaria diagnosis," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2016), pp. 493–496.
26. T. Zhou, H. Lu, Z. Yang, *et al.*, "The ensemble deep learning model for novel COVID-19 on CT images," Appl Soft Comput **98**, 106885 (2021).
27. A. Sehgal and N. Kehtarnavaz, "Guidelines and Benchmarks for Deployment of Deep Learning Models on Smartphones as Real-Time Apps," Mach Learn Knowl Extr **1**(1), 450–465 (2019).
28. H. Wang, L. Barriga, A. Vahidi, *et al.*, "Machine Learning for Security at the IoT Edge - A Feasibility Study," in *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW)* (2019), pp. 7–12.
29. W. Dai and D. Berleant, "Benchmarking Contemporary Deep Learning Hardware and Frameworks: A Survey of Qualitative Metrics," in *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)* (2019), pp. 148–155.