


Article

Underwater Object Detection and Reconstruction Based on Active Single-Pixel Imaging and Super-Resolution Convolutional Neural Network

Mengdi Li ^{1,2}, Anumol Mathai ², Stephen L. H. Lau ², Jian Wei Yam ², Xiping Xu ^{1,*} and Xin Wang ² 

¹ College of Optoelectronic Engineering, Changchun University of Science and Technology, Changchun 130022, China; 2017100212@mails.cust.edu.cn

² School of Engineering, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway 47500, Malaysia; anumol.mathai@monash.edu (A.M.); stephen.lau@monash.edu (S.L.H.L.); Jian.yam@monash.edu (J.W.Y.); wang.xin@monash.edu (X.W.)

* Correspondence: xxp@cust.edu.cn; Tel.: +86-182-4307-0831

Abstract: Due to medium scattering, absorption, and complex light interactions, capturing objects from the underwater environment has always been a difficult task. Single-pixel imaging (SPI) is an efficient imaging approach that can obtain spatial object information under low-light conditions. In this paper, we propose a single-pixel object inspection system for the underwater environment based on compressive sensing super-resolution convolutional neural network (CS-SRCNN). With the CS-SRCNN algorithm, image reconstruction can be achieved with 30% of the total pixels in the image. We also investigate the impact of compression ratios on underwater object SPI reconstruction performance. In addition, we analyzed the effect of peak signal to noise ratio (PSNR) and structural similarity index (SSIM) to determine the image quality of the reconstructed image. Our work is compared to the SPI system and SRCNN method to demonstrate its efficiency in capturing object results from an underwater environment. The PSNR and SSIM of the proposed method have increased to 35.44% and 73.07%, respectively. This work provides new insight into SPI applications and creates a better alternative for underwater optical object imaging to achieve good quality.

Keywords: single-pixel imaging; compressive sensing; super-resolution convolutional neural network



Citation: Li, M.; Mathai, A.; Lau, S.L.H.; Yam, J.W.; Xu, X.; Wang, X. Underwater Object Detection and Reconstruction Based on Active Single-Pixel Imaging and Super-Resolution Convolutional Neural Network. *Sensors* **2021**, *21*, 313. <https://doi.org/10.3390/s21010313>

Received: 23 November 2020

Accepted: 29 December 2020

Published: 5 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the underwater environment, reconstruction of objects is a challenging task, due to light attenuation by water absorption and the illumination volatility by the scattering medium [1,2]. Various approaches have been developed for imaging objects under low-light conditions and for improving the quality of underwater images. Polarimetric imaging systems [3,4] were used to reduce the backscattering effect and produced a good quality image. Although polarization correction can improve the renovation quality, it also blocks part of the light incident on the detector, thereby reducing the signal to noise ratio and complicating the problem of imaging underwater. Range-gated imaging was developed in Mariani et al. [5] by combining the time of flight technique. The use of a pulsed laser as a light source enabled the authors to measure the distance of the object from the light source, thereby calculating the depth of the object by eliminating the backscattering effect. However, relative to other cameras, range-gated systems are generally more expensive, more complicated to operate, and limited in resolution and frame rate. Geotagging and color correction approaches were implemented to successfully generate 2D and 3D images of the underwater environment [6]. To capture underwater images, a digital camera with waterproof housing was used and the location of the camera was identified through geo-tagging. Though it could produce good-quality 2D and 3D maps of the underwater environment, requirements such as clear water and calm surface conditions limited its

performance. To further improve the underwater image quality, in Lu et al. [7] used a multi-scale cycle generative adversarial network. In that, the discriminator could produce high-quality images under homogeneous lighting conditions. However, it is failed under inhomogeneous illumination. The work was extended in Tang et al. [8] using an algorithm called dark channel prior to further improving underwater image quality. Empirical mode decomposition was developed in Çelebi et al. [9], where obtained underwater images were decomposed based on the spectral components and then, again, constructed by combining the intrinsic mode functions to enhance visual image quality. However, this three-channel component calculation method increased the amount of calculation. Most of the schemes utilize the silicon-based Charge Coupled Device (CCD) or Complementary Metal Oxide Semiconductor (CMOS) to image underwater objects. The limited response of the detection medium to a specific bandwidth has always been a drawback of those sensors imaging systems under low illumination conditions.

Recently, single-pixel imaging (SPI) has attracted widespread attention in imaging objects under low-light conditions and a scattering medium. SPI uses pre-programmed modulated light patterns and knowledge of the scene under view to acquire spatial information on the target [10,11]. In projection-based SPI, the object is illuminated with two-dimensional spatially coded patterns and collects reflected or transmitted light signals from the object with the single-pixel detector (SPD) to gather the fine details of the object [11]. In addition to cost-effectiveness, other advantages of SPI are low dark current and light sensitivity, producing good-quality images. Simultaneously, the introduction of compressed sensing (CS) in SPI has enabled reconstruction with fewer measurements [12,13]. Owing to these improvements, SPI has been exploited in scattering media [14,15]. Although it has been proved that SPI can be imaged underwater, the fluctuation in illumination caused by absorption and the scattering medium of water still seriously reduce the reconstruction quality. To suppress these effects. Chen, et al. [16] proposed a SPI detection scheme in which an object can be recovered through turbid water by transmission signal data. Although it has been proved that SPI can be imaged underwater, the scattering and absorption of water seriously reduce the reconstruction quality. In addition, these methods also have some noise on the detector, decreasing the image quality [17]. Hence, for SPI underwater, an SPI restoration algorithm that is robust to noise is demanded. Moreover, limitations in the visual aspect of the above imaging system can be overcome by combining both deep learning and a basic SPI system.

Deep learning [18], as an emerging field of research, has proven its performance in image processing, such as image recognition [19,20], object detection [21], and person pose estimation [22]. When combining neural networks with SPI systems, there are two ways to get the final target image. One exploits the robust neural network to rehabilitate the object image and the other trains and predicts images with the neural network after reconstruction. For the former, Higham, et al. [23] utilized MATCONVNET to create a deep convolutional auto-encoder using the patterns as the encoding layer and Hadamard patterns as the optimization basis. For subsequent decoding layers, three conventional layers were used to recover the real-time high-resolution video. For the latter scheme, Rizvi, et al. [24] raised a deep convolutional autoencoder network that uses reconstructed under-sampled training and testing images as input and high-quality predicted images as output. Compared with previous methods, deep learning algorithms have been proved to be more reliable for restoration. Therefore, the employment of deep learning in SPI system has brought drastic changes in image quality.

Though different studies about the influence of turbulence [25] and deep learning on SPI are discussed in Dutta et al. [26] and Jauregui-Sánchez et al. [27], there are limited studies that focus on SPI with deep learning for underwater image reconstruction. Therefore, in this paper, we leverage single-pixel imaging and deep learning to improve underwater imaging restoration. An experimental setup was established to verify the presented method. The effectiveness of our technique is demonstrated through both experiments and simulation. We summarize them as follows:

- An optical imaging system based on SPI is developed for imaging objects in the underwater environment. Our experimental results validated that the recovered object by the underwater SPI system is affected by scattering and absorptions.
- Compressive sensing-based super-resolution convolutional neural network (CS-SRCNN) is implemented by combining the advantages of the SRCNN and SPI system. The newly introduced CS-SRCNN takes reconstructed underwater SPI images to train the network and to predict high-resolution images.
- We also demonstrate the effectiveness of our technique through simulation. In the simulation, the proposed method can restore objects with a low sampling rate and can produce more robust reconstructions. Our experimental results also validated that the recovered object by the underwater SPI system is affected by scattering and absorptions.
- We experimentally demonstrated reconstruction to get better results with a low sampling rate of only 30%.

2. Underwater Object Reconstruction

2.1. Theory

2.1.1. Compressive Sensing

Figure 1 shows the schematic diagram of the SPI system. Assuming that X is the column vector reshaped from the $N = p \times q$ pixel image and that it is sparsely sampled by measurement matrix Φ , the corresponding measurement data can be expressed as follows:

$$Y = \Phi X + e \quad (1)$$

where Y is the measurement data by stacking of y_i , which is an $M \times 1$ column vector having linear measurements. M is the number of measurements used for image acquisition. The measurement matrix $\Phi \in \mathbb{R}^{M \times N}$ contains M row vectors, which are the stacking of the 2-dimension coded patterns ϕ_m . e of size $M \times 1$ denotes the noise. Equation (1) can be written as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,N} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{M,1} & \phi_{M,2} & \cdots & \phi_{M,N} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix} \quad (2)$$

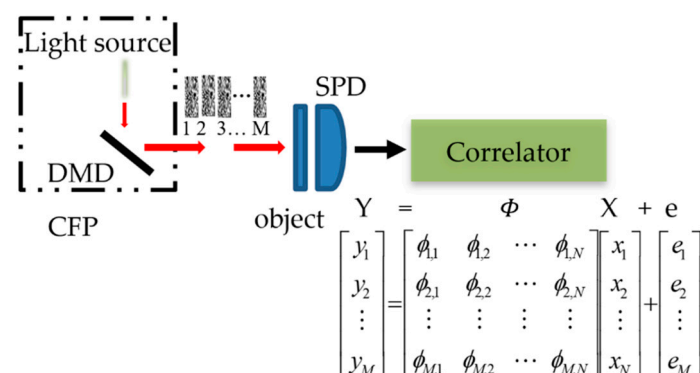


Figure 1. The single-pixel imaging setup. DMD: Digital Micromirror Device, CFP: Calculated Filed Pattern, and SPD: Single-Pixel Detector.

A measurement matrix Φ was constructed with a random pattern combined with the observed signal y_i to obtain a system of equations. However, as mentioned above, Φ has fewer rows than columns ($N = p \times q$). Therefore, solving out X using Y and Φ seems impossible because the solution of the system of equations is not unique. To

address the issue, we relied on the principle of compressed sensing (CS) [28], which can be used effectively for sparse images. Sparsity indicates that a signal is represented in an appropriate basis, where most of the information in images are close to zero. In most instances, the gradient integral of a natural signal is statistically low; hence the image can be compressed by the CS method. When X is k -sparse, it can be recovered with $M = O(k \log N)$ incoherent nonadaptive linear measurements. Equation (1) can be summarized as follows:

Let $D_i x$ denote the discrete gradient vector of X at position (i, j) , D be the gradient (horizontal $D_i^h x$ and vertical $D_j^v x$) operator with the following operational definitions:

$$D_i^h x = \begin{cases} x_{i+1,j} - x_{i,j}, & 1 \leq i < M \\ 0, & i = M \end{cases} \quad (3)$$

$$D_j^v x = \begin{cases} x_{i,j+1} - x_{i,j}, & 1 \leq j < N \\ 0, & j = N \end{cases} \quad (4)$$

The total variation regularization (TV) in X is simply the sum of the magnitudes of this discrete gradient at each point and the error term:

$$\|x\|_{TV} = \sum_{i,j} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2} \quad (5)$$

$$TV(X) = \sum_i \|D_i x\|_1 + \frac{\mu}{2} \|Y - \Phi X\|_2^2 \quad (6)$$

where $\sum_i \|D_i x\|_1$ is the discrete TV of X , μ is a constant scalar used to balance these two terms, and $\|x\|_1 = \sum_{i=1}^N |x_i|^1$ represents ℓ_1 norm. TV regularization to recover the image is discussed in Li [29]. The first term in Equation (6) is small when $D_i x$ is sparse. When the optimal X is consistent with Equation (1) with a small error, the second term is small [30]. The image reconstruction problem of Equation (1) can be expressed as follows:

$$\min TV(X) \text{ subject to } \Phi X = Y \quad (7)$$

Accurate recovery can be achieved by solving a convex optimization program that is easy to handle [31].

2.1.2. The Super-Resolution Convolutional Neural Network

Convolutional neural network (CNN), a typical deep learning algorithm, has been widely used in image processing due to its powerful feature learning function in computer vision research. SRCNN [32,33] is the first end-to-end super-resolution algorithm using the CNN structure. An SRCNN architecture was adopted as the benchmark, which learns to map between low or high-resolution images. The low-resolution (LR) components, X_{SPI} , were loaded into bicubic interpolation to get interpolated components X_{II} . The patch extraction layer extracted patches from the bicubic interpolated components, and feature extraction by convolution can be denoted as follows:

$$F_1 = \max(0, W_1 * X_{II} + B_1) \quad (8)$$

where the variables F , X_{II} , W_1 , and B_1 represent the mapping function, the original high-resolution (HR) image after interpolation, the filters, and the biases (n_1 -dimensional vector), respectively, and $*$ represents the convolution operation. The size of W_1 is $n_1 \times c \times f_1 \times f_1$, f is the size of the filter, c is the number of channels contained in the input image, and n is the number of convolution kernels, with suffix 1 indicating the first layer. A Rectified Linear Unit (ReLU) was used as the activation function of the network. The function expression of ReLU is as follows:

$$\text{ReLU} = \max(0, x) \quad (9)$$

ReLU can be computed faster than traditional activation functions such as sigmoid and allows for easier optimization of the neural network. The nonlinear mapping layer maps from LR space to HR space. The operation is as follows:

$$F_2 = \max(0, W_2 * F_1 + B_2) \quad (10)$$

The size of W_2 is $n_2 \times n_1 \times f_2 \times f_2$, with B_2 being an n_2 -dimensional vector. This layer outputs the n_2 -dimensional feature map as an input to the third layer.

The reconstruction of HR images in the third layer can be expressed as follows:

$$F_3 = \max(0, W_3 * F_2 + B_3) \quad (11)$$

where the size of W_3 is $c \times n_2 \times f_3 \times f_3$. Combining these three operations constitutes a CNN in which all filter weights and deviations are optimized. In addition, during the training phase, the mapping function F needs to estimate the network parameters $\Theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$. The reconstructed images were described as $F(X_{II}; \Theta)$, and the HR image was X_{SR} . The error between $F(X_{II}; \Theta)$ and the real image X_{SR} was minimized. To achieve this goal, the mean square error (MSE) was chosen to construct the loss function of the network model of this paper. The loss function L is expressed as follows:

$$L(\Theta) = \frac{1}{o} \sum_{i=1}^o \|F(X_{IIi}; \Theta) - X_{SRi}\|^2 \quad (12)$$

where o is the number of training images and X_{SRi} is a set of HR image. The loss function was minimized by applying the gradient descent method and the standard back-propagation algorithm.

2.2. The Reconstruction of Underwater Single-Pixel Imaging based on Compressive Sensing and Super-Resolution Convolutional Neural Network

According to those theories, we regarded CS image regeneration as an inverse problem and tackled this question based on SRCNN. In this study, we propose an underwater object reconstruction CS-SRCNN method which combines the improved SRCNN and SPI system. In addition, the specific implementation architecture of the proposed method is shown in Figure 2. This method architecture was adopted as the benchmark, which learns to map between low or high-resolution images. That is, high-frequency components were increased in low-resolution (LR) components. In the mapping process, network measurement data were taken as input and produce high-resolution (HR) images were taken as output [34,35].

The architecture combined both deep learning and traditional sparse coding technique (in Figure 3). The CS-SRCNN can be divided into their respective functions, namely patch extraction, nonlinear mapping, and final reconstruction. The first part of the SRCNN (patch extraction) consisted of a CS layer and a three-layer CNN. The input to the entire neural network was M -dimensional compressed raw data from SPI system. The raw data ran iterative processing by TV regularization to attain LR components. The network enlarged the LR components by bicubic interpolation, which is a preprocessing step for CNN. Compared to other methods, bicubic interpolation is faster and does not introduce too much additional information. After that, the first convolutional layer extracted 9×9 LR patches. The output was passed to the rectified linear unit (ReLU) activation function. The second part of the SRCNN (nonlinear mapping) made use of a convolution layer with a kernel size of 1×1 to output 64 feature maps, which were then concatenated into a matrix. The output was passed to the ReLU activation function. In the third part of the SRCNN (reconstruction), the matrix from the preceding layer was passed through a convolution layer with a kernel size of 5×5 . HR feature vectors were reconstructed, and all HR patches were combined to form the highest enough HR components. Subsequently, we applied postprocessing (e.g., filtering to remove noise) to the reconstructed images.

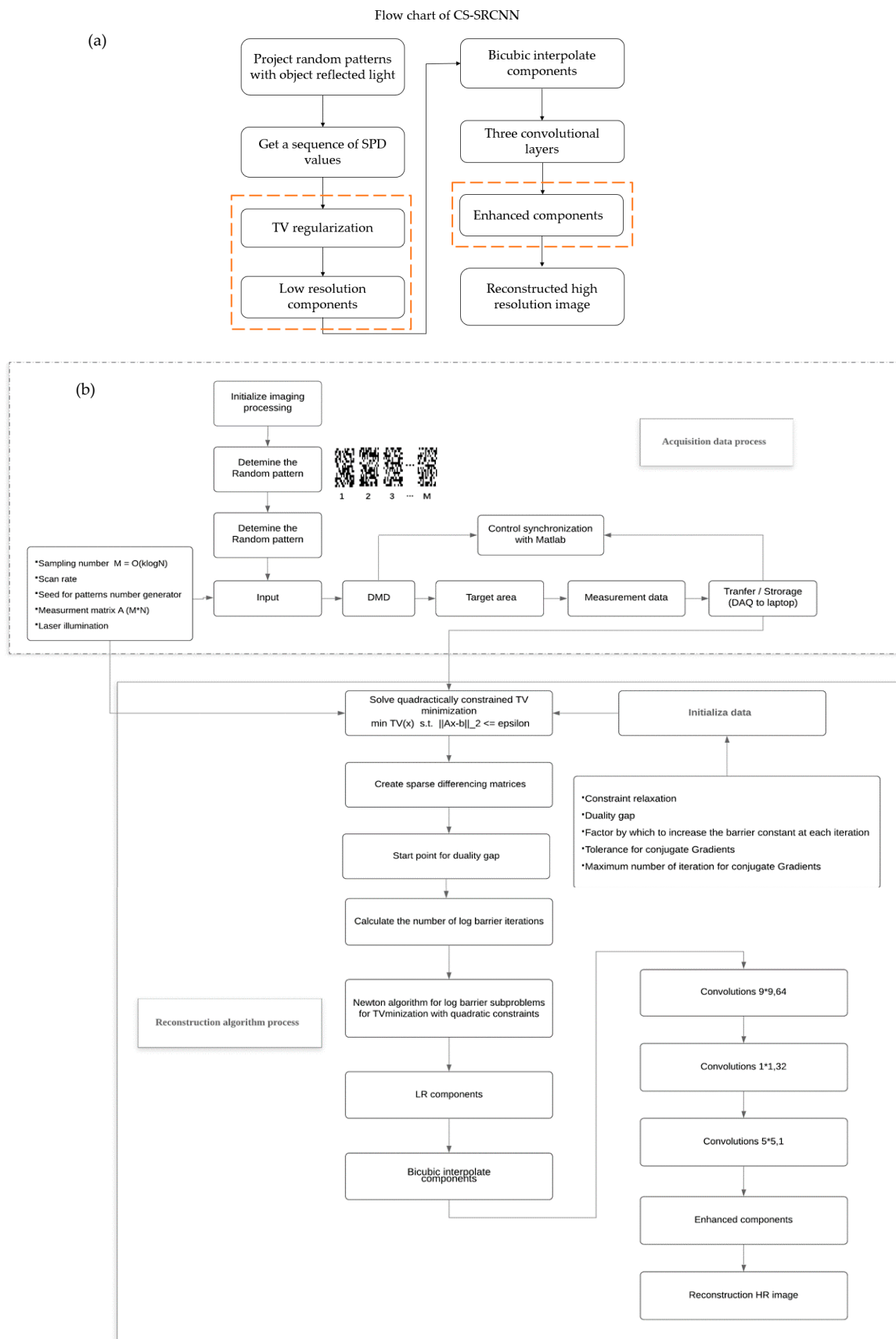


Figure 2. Activity diagram of the proposed reconstruction algorithm: (a) traditional single-pixel imaging and super-resolution convolutional neural network (SRCNN), and (b) the proposed method in this paper.

In our study, the dataset was divided into two subsets: the training set and test set. We used 400 images of handwritten digits downloaded from the MNIST database with the corresponding SPI images to train the SRCNN. The other 100 images were put into the test set. MNIST handwritten digits (28×28 resolution) were resized into the correct resolution (e.g., 32×32). For each training image, the same set of M different random patterns was applied in both the simulation and experiment (more details in Section 3). After the set of M patterns was run with the object, we attained a light intensity signal y_i of length M . We then paired up the resulting signal y_i and the corresponding 2D image. The objects were replaced, and the above simulation process was rerun to obtain another pair of labeled data. For training, no additional noise was added to the simulated light intensity signal. This process was run repeatedly across the whole 400 images in the training set and were paired up with y_i .

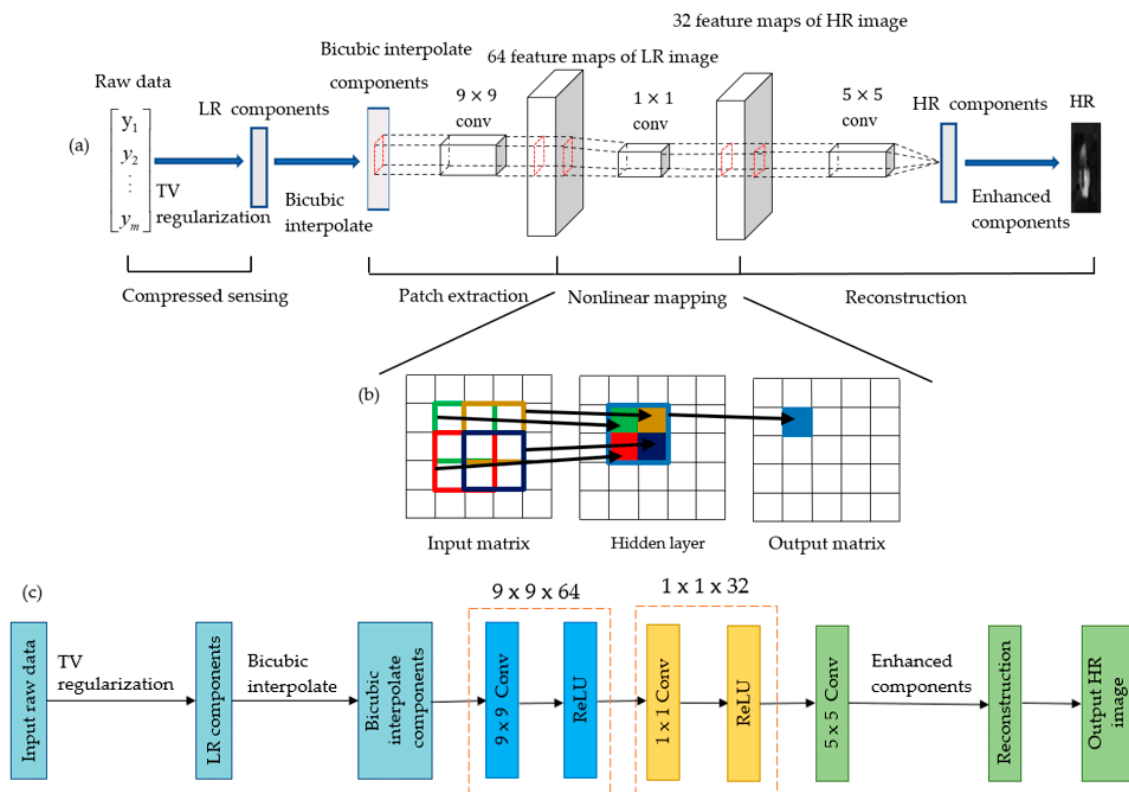


Figure 3. Architecture of the proposed reconstruction: (a) architecture of the proposed reconstruction; (b) convolved data from the input matrix to the output matrix, where the convolution process moves 1 step in the input matrix; and (c) a schematic diagram of our method.

3. Simulation Results and Analysis

In the underwater SPI system, to investigate the effect of the number of measurements and water turbidity, random patterns were employed to sample the object with 32×32 resolution. The simulation method using Gaussian blur [36] (different Gaussian noises of 10, 15, and 25 dB were added to simulate different underwater turbidity conditions) as the disturbance factor simulated scattering of the water medium and qualitatively analyzed the non-disturbance ability of SPI under different sampling rates in turbid water. The simulation results for different sampling rates are shown in Figure 4. We can see that there are blurry artifacts in the refurbished images at different sampling rates between 4.8–30%.

Furthermore, to inspect how the image quality changes in the underwater environment, we simulated CS-SRCNN reconstruction for an image of the letter G with a sampling rate of 0.29 for different underwater turbidities. We then verified the influence of turbid-

ity on the reconstruction quality by applying the conventional SRCNN and our method. According to our setup, we compared images from SPI, SRCNN, and our method by applying upscaling factors of 3 on the underwater image. Figure 5 shows the visual effects at different turbidities (0, 20, 40, and 60 nephelometry turbidity unit (NTU)). As the turbidity increased from 0 NTU to 60 NTU, we observed a significant degradation of the reconstructed image, comparing Figure 5b with Figure 5e–g. However, CS-SRCNN can reconstruct clear images, even if the turbidity is as high as 60 NTU, and the reconstructed image can still be satisfactory with a sampling ratio of about 29%.

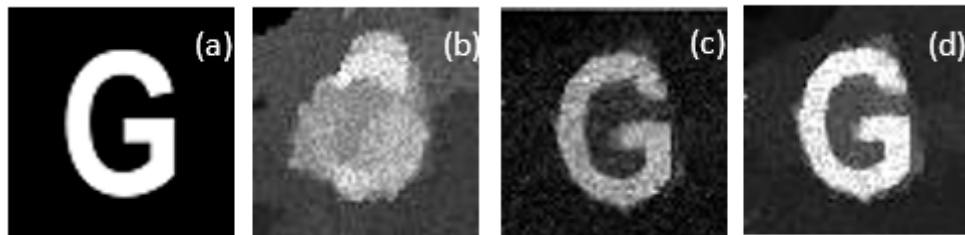


Figure 4. The images of different compression ratios that were reconstructed: (a) an original target image with pixels size 32×32 and (b–d) reconstructed images using different numbers of random speckle patterns (the M takes the values 50, 200, and 300, respectively, and different compression ratios from 4.88% and 19.53% to 29.29%).

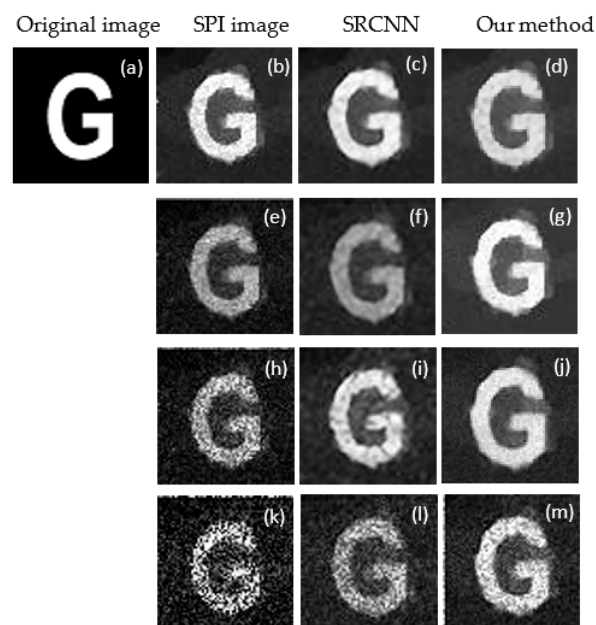


Figure 5. Image reconstruction results at different turbidities using random patterns in clear water under 300 measurements: (a) original image; (b–d) single-pixel imaging (SPI) image, SRCNN image, and our method image with 0 NTU; (e–f) SPI image, SRCNN image, and our method image with 20 NTU; (g) SPI image, SRCNN image, and our method image with 40 NTU; and (k–m) SPI image, SRCNN image, and our method image with 60 NTU.

Figure 6 compares the peak signal to noise ratio (PSNR) and structural similarity index (SSIM) of images taken with different turbidities of SPI before reconstructed with TV-regularization, SRCNN, and CS-SRCNN. We kept the number of measurements constant at 300. The figure shows that both PSNR and SSIM decreased with turbidity for all three methods of reconstruction. In the low turbidity region from 0 NTU to 40 NTU, the PSNR and SSIM of the three methods declined at a similar rate. On the contrary, in areas where the turbidity is higher than 40 NTU, the imaging quality of our method shows robustness (the slope tends to be flat) while the other two approaches followed the former trend. Since these three evaluation factors have good robustness to the influence of the scattering

factor from the water medium, this makes CS-SRCNN more sensitive, and the results of underwater imaging can be further enhanced.

We also compared the image recovered by our method with the images reconstructed by the end-to-end learning method and image processing method, which is the scattering media image recovery method based on a polarimetric ghost imaging system formulated by Li, et al. [37]. The PSNR and SSIM were used to evaluate and compare the results. The results of those methods are shown in Table 1. Compared with Li et al.'s method, the reconstructed image PSNRs and SSIMs in our method were significantly increased by 17.39%/83.78% and 16.11%/90.32%, which means that our method has better performance than the other methods.

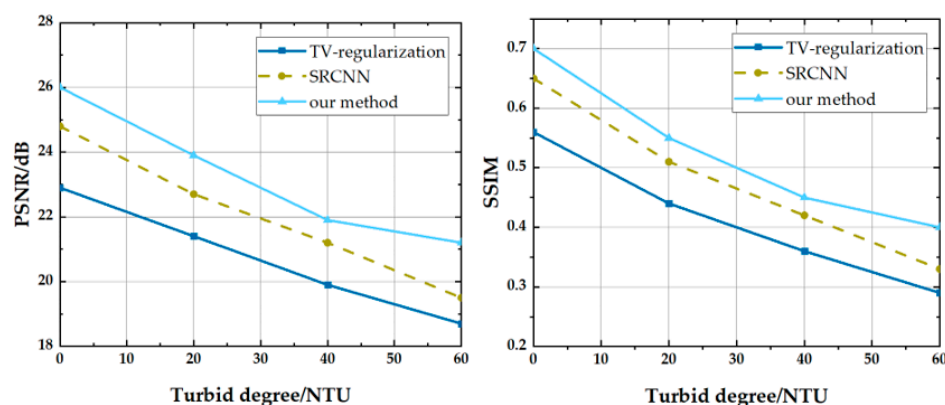


Figure 6. The index values for Bicubic and SRCNN at different magnifications.

Table 1. Comparative analysis of different recovery methods with a sampling rate of 18.3% and scattering condition (the signal to noise ratio is 10 dB).

	"Lena"				"Cameraman"			
	SRCNN	CS	Li	Our	SRCNN	CS	Li	Our
PSNR	21.92	18.83	20.47	24.03	21.83	18.48	20.30	23.57
SSIM	0.56	0.32	0.37	0.68	0.57	0.30	0.31	0.59

4. Experimental Results and Analysis

4.1. Experimental Setup

The experimental setup for the underwater SPI system is illustrated in Figure 7, which consists of a laser source and a DMD to project binary patterns. A fast response time SPD with a collection lens was used to measure the transmission intensity resulting from each pattern. The corresponding intensity values were captured by a data acquisition (DAQ) device. The object image was acquired by correlating the values from the DAQ with the pattern matrix.

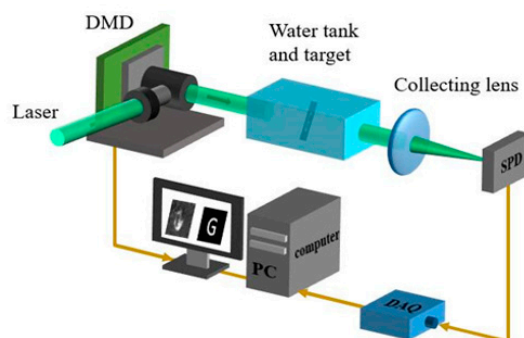


Figure 7. Schematic diagram of the underwater object detection system.

The experimental setup is shown in Figure 8. The continuum laser is monochromatic, highly intense, and power-efficient and produces a coherent wavelength of 520 nm (the laser power was 20 mW). It is, therefore, a viable choice of light source that illuminates the DMD. The DMD allows for a configurable input (MATLAB programming and light source)/output (random pattern) trigger for convenient synchronization with continuum laser and SPD peripheral devices. The DLP (DLP Light Crafter 6500, Texas Instruments) provides a true HD resolution of 1920×1080 , and more than 2 million programmable micromirrors by the DMD chip mounted on it. The pre-programmed random patterns are produced by the DMD, which is in the form of a 32×32 matrix. Each light pattern that passed through the underwater object (transparent “G”; the length, width, and height of the water tank were 20 cm, 20 cm, and 25 cm) was directed to a collecting lens (the model used was THORLABS LMR75/M). The collecting lens focused the light on the SPD (the model used was PDA36A2), which acted as a bucket detector and had a bandwidth of 12 MHz and an area of $13 \times 13 \text{ mm}^2$. The total light intensity value detected by the digital-to-analog converter (National Instrument DAQ USB-6001) was converted to an electrical signal and was transmitted to a laptop via USB for image reconstruction. The synchronization between DMD and SPD was done by sending a trigger signal to the DMD to refresh the patterns. Simultaneously, the trigger signal was sent to the DAQ to attain the SPD voltage. The data management and sampling were controlled and synchronized by MATLAB software. The automation scripts for data collection were executed with the MATLAB software on a laptop. After minimizing the total variation, the reconstructed algorithm was implemented to reconstruct the image.

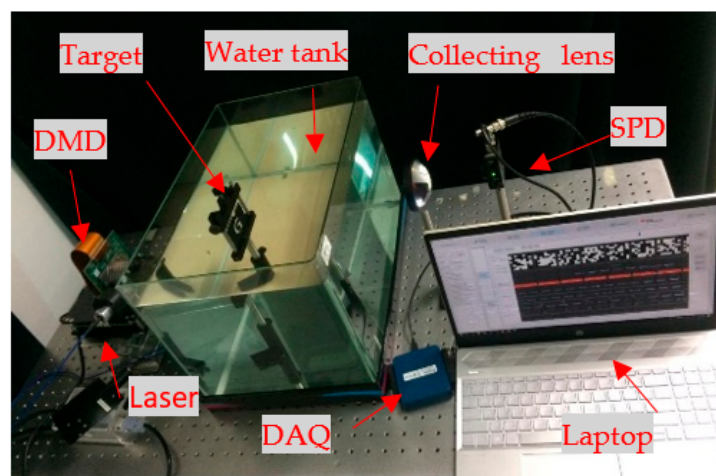


Figure 8. Experimental setup implemented in our lab environment.

4.2. TV Regularization (Compressive Sensing) Reconstruction

The experimental results of the reconstruction of underwater SPI using TV regularization are shown in Figure 9. The reconstruction was affected by the number of measurements, imaging speed, and compressibility. Random patterns were employed to reconstruct the object with 32×32 resolution. To investigate the effect of the number of measurements on the underwater image, different numbers of measurements were used to capture and reconstruct the images.

Also, Figure 9 represents the reconstruction results for three different numbers of random patterns. The images in Figure 9b–d,f–h represent the reconstruction results for 100, 200, and 300 random patterns, respectively. It can be seen that the simple object (“+”) was restored better than the complex object (“G”). Furthermore, the object can be reconstructed even with 29% of the total number of pixels in the original image by using CS.

In order to quantify the accuracy of image reconstruction results, image quality assessment techniques such as the mean square error (MSE), PSNR, SSIM, and visibility

(V) were employed. The MSE was calculated by comparing the reconstructed image and original object image, and it is defined as follows:

$$MSE = \frac{1}{pq} \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} [x(i, j) - X_R(i, j)]^2 \quad (13)$$

where x is the original image with $p \times q$ resolution, and X_R is the reconstructed image. Suppose $x_{\max}^2 = 2^K - 1$, where K represents the number of bits used for a pixel, $K = 8$, and $x_{\max} = 225$. Similarly, PSNR describes the ratio of original image pixels to reconstructed pixels, and it is defined as follows:

$$PSNR(dB) = 10 \lg \frac{x_{\max}^2}{MSE} = 10 \lg \left[\frac{x_{\max}^2 pq}{\sum_{i=0}^{p-1} \sum_{j=0}^{q-1} [x(i, j) - X_R(i, j)]^2} \right] (dB) \quad (14)$$

SSIM is a full-reference metric that describes the statistical similarity between two images. The indicator was first proposed by the University of Texas at Austin's Laboratory for Image and Video Engineering [38]. This is given by the following:

$$SSIM = \frac{(2\mu_x \mu_{X_R} + c_1)(2\sigma_{xX_R} + c_2)}{(\mu_x^2 + \mu_{X_R}^2 + c_1)(\sigma_x^2 + \sigma_{X_R}^2 + c_2)} \quad (15)$$

where μ_x and μ_{X_R} are the averages of x and X_R , respectively; σ_x and σ_{X_R} are standard deviations of x ; σ_{xX_R} are covariances of x and X_R , respectively; c_1 and c_2 are variables to stabilize the division with the weak denominator (constants); $c_1 = (K_1 L)^2$; and $c_2 = (K_2 L)^2$. Generally, $K_1 = 0.01$, $K_2 = 0.03$, and $L = 255$ (L is the dynamic range of the pixel value, generally taken as 255). The visibility (V) is defined as follows:

$$V = \frac{\langle S_{in} \rangle - \langle S_{out} \rangle}{\langle S_{in} \rangle + \langle S_{out} \rangle} \quad (16)$$

where $\langle S_{in} \rangle$ and $\langle S_{out} \rangle$ are the average values of SPI of the interesting region and the background region, respectively. The PSNR, SSIM, and V were calculated using the MATLAB function.

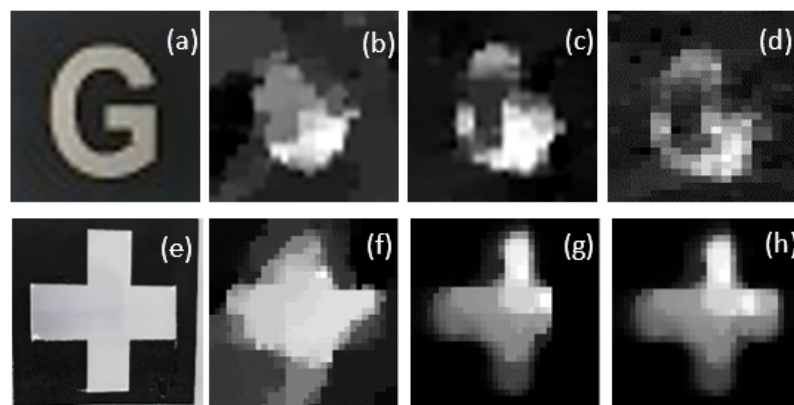


Figure 9. Images of different compression ratios reconstructed: (a) an original target image with pixel size 32×32 ; (b–d) reconstructed images using different numbers of random speckle patterns (M takes the values 100, 200, and 300, respectively, and different compression ratios from 9.76% and 19.53% to 29.29%); (e) an original target image with pixel size 32×32 ; and (f–h) reconstructed images using different numbers of random speckle patterns.

As can be seen from Tables 2 and 3, the measurements are directly proportional to the PSNR and SSIM. To be clear, the PSNR and SSIM ranges are higher for image restoration at more measurements. For example, in Table 2, PSNR = 11.59 dB, SSIM = 0.26, and V = 0.29

for 300 measurements and PSNR = 9.02dB, SSIM = 0.10, and V = 0.24 for 100 measurements. Compared with the third column, the reconstructed images PSNR, SSIM, and V in the first column were significantly increased by 28.49%/1.6/20.83%. Therefore, the more the number of measurements, the better the similarity of images reconstruction.

Table 2. The respective peak signal to noise ratio (PSNR) and structural similarity index (SSIM) for different measurements (letter “G”).

	M = 100	M = 200	M = 300
PSNR	9.02	10.17	11.59
SSIM	0.10	0.13	0.26
V	0.24	0.27	0.29

Table 3. The respective PSNR and SSIM for different measurements (object “+”).

	M = 100	M = 200	M = 300
PSNR	10.25	11.58	12.78
SSIM	0.23	0.31	0.42
V	0.27	0.29	0.31

4.3. Reconstruction Based on Compressive Sensing Super-resolution Convolutional Network (CS-SRCNN)

In this scheme, the data set with 300 measurements from an underwater SPI system was fed into the TV regularization algorithm. The resultant output enlarged the desired matrix size by bicubic interpolation. The interpolated components were set as the input to the following layers. The output of each layer of convolution is shown in Figure 10. For the first layer, the convolution kernel size was 9×9 ($f_1 \times f_1$), the number of convolution kernels was $n_1 = 64$, and the output was 64 feature maps. Optimization of the output was performed by ReLU as a nonlinear activation function. For the second layer, the convolution kernel size was 1×1 ($f_2 \times f_2$), the number of convolution kernels was $n_2 = 32$, and the output was 32 feature maps. Similarly, for the third layer, the convolution kernel size was 5×5 ($f_3 \times f_3$), the number of convolution kernels was $n_3 = 1$, and the output was 1 feature map. The final feature map is the reconstructed HR components.

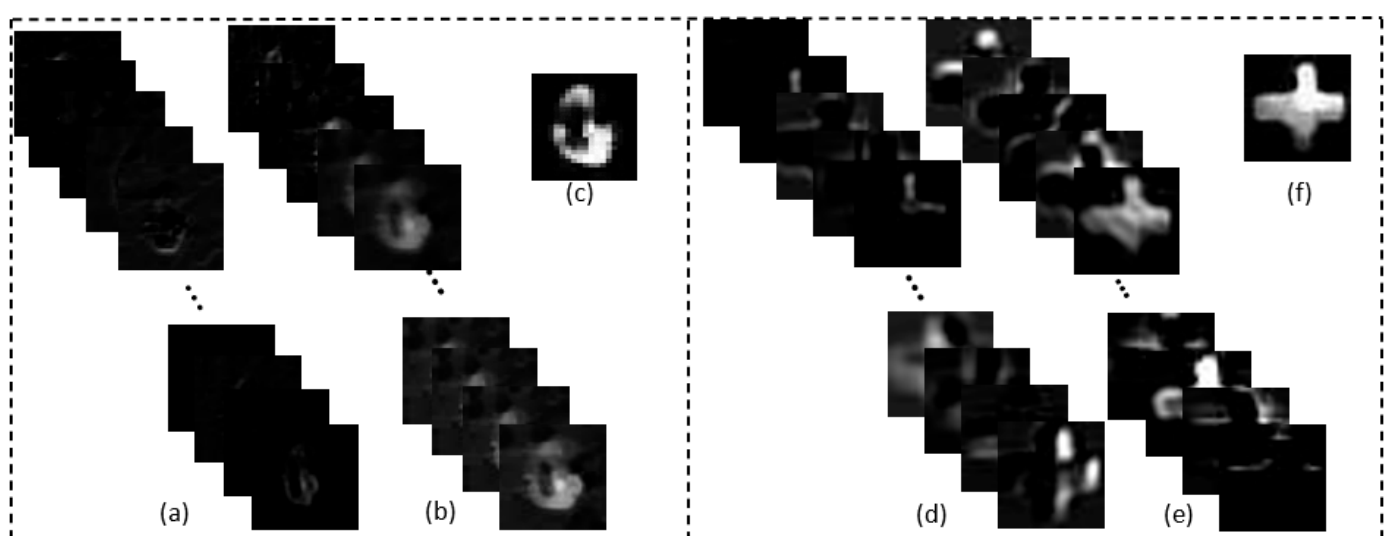


Figure 10. The results of each layer of convolution: (a,d) part of the first 64 feature maps, (b,e) part of the second layer of 32 feature maps, and (c,f) part of the 1-feature map of the third layer.

Furthermore, in this part, according to our setup, a comparison was made for SPI, SRCNN, and our method images by applying upscaling factors of 3 on the underwater image. Figure 11 shows the visual effects of different methods.

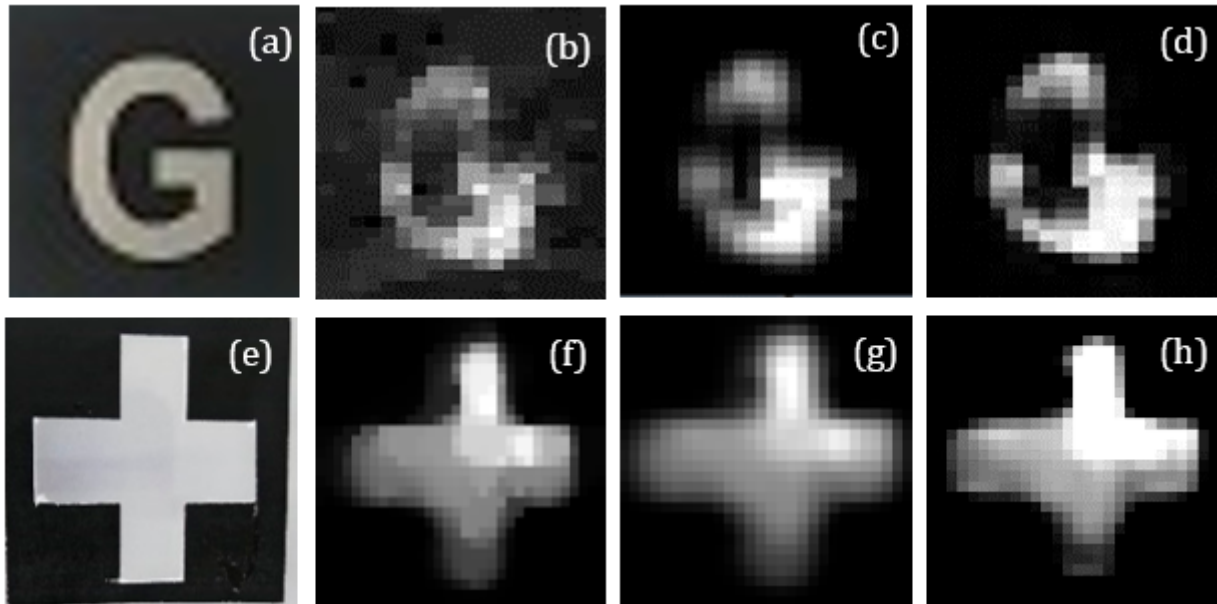


Figure 11. Image reconstruction results at different methods using random patterns in clear water under 300 measurements: (a,e) original image, (b,f) SPI image, (c,g) SRCNN image, and (d,h) our method image.

To investigate the relationship among the image quality of super-resolution images reconstructed by the CS-SRCNN scheme and SPI image, two indicators are commonly used such as PSNR and SSIM. The analysis of results focused more on the complex letter “G”. The higher PSNR and SSIM, the closer the pixel value to the standard. With the idea of controlling the variates, the index values for SRCNN and our method at upscaling factors of 3 are presented in Tables 4 and 5 by keeping other conditions such as compression ratio of the reconstructed image, the reconstructed image algorithm, and so on constant. Moreover, the respective PSNR and SSIM for each recovered image in Figure 9d,h are shown in Tables 4 and 5. Compared with the third column, the reconstructed images PSNR, SSIM, and V in the second column increased significantly by 29.83%/40.62%/10.00%, and 15.97%/44.00%/6.45%, respectively. We can know that PSNR, SSIM, and V have been significantly improved, consistent with imaging theory. The results validate our analysis of neural networks in underwater SPI.

Table 4. The respective PSNR and SSIM for different methods (the letter “G”).

	SPI Image	SRCNN Image	Our Method Image
PSNR	11.59	12.07	15.67
SSIM	0.26	0.32	0.45
V	0.29	0.30	0.33

Table 5. The respective PSNR and SSIM for different methods (object “+”).

	SPI Image	SRCNN Image	Our Method Image
PSNR	12.89	13.21	15.32
SSIM	0.23	0.25	0.36
V	0.31	0.31	0.33

The above analysis only evaluates the quality of images from the two aspects PSNR and SSIM. To comprehensively evaluate the performance of the algorithm, a comparative analysis was performed using the original image shown in Figure 12a at different compression ratios for TV regularization and our algorithms.

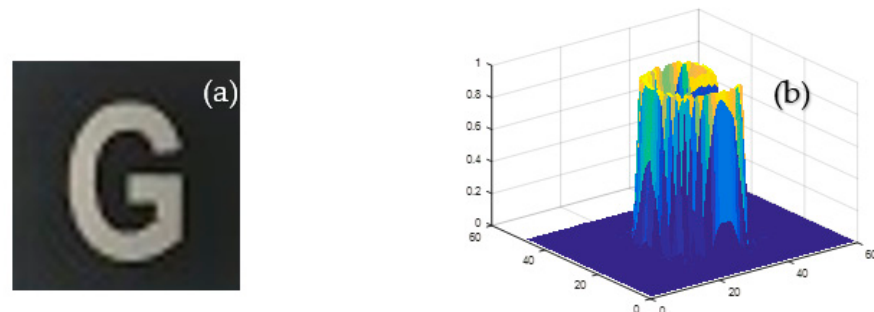


Figure 12. Test image: (a) original image and (b) the three-dimensional surface of the original image.

In comparison, the images with compression ratios of 0.70, 0.13, and 0.20 were reconstructed. Figure 13a–c shows the reconstruction results for TV regularization for different compression ratios. From the figure, the signal reconstructed is closer to the original image signal when compression ratio increases and convergence occurs. Figure 13d–f are the recovery results corresponding to our reconstruction algorithm. As the compression rate increases, the reconstruction accuracy of the image gets higher and higher; however, the result shows that there are peaks, which are quite different from the original signal distribution curve. The main reason for this is that our reconstruction algorithm performs block processing on the input image and then reconstructs each small block independently and combines it into the whole picture. Therefore, there is an image blocking artifact problem which is not obvious when the compression ratio is high but is more evident at a low compression ratio. Filtering and smoothing can be performed at the end of reconstruction to alleviate the blocking artifact problem in the resultant HR image.

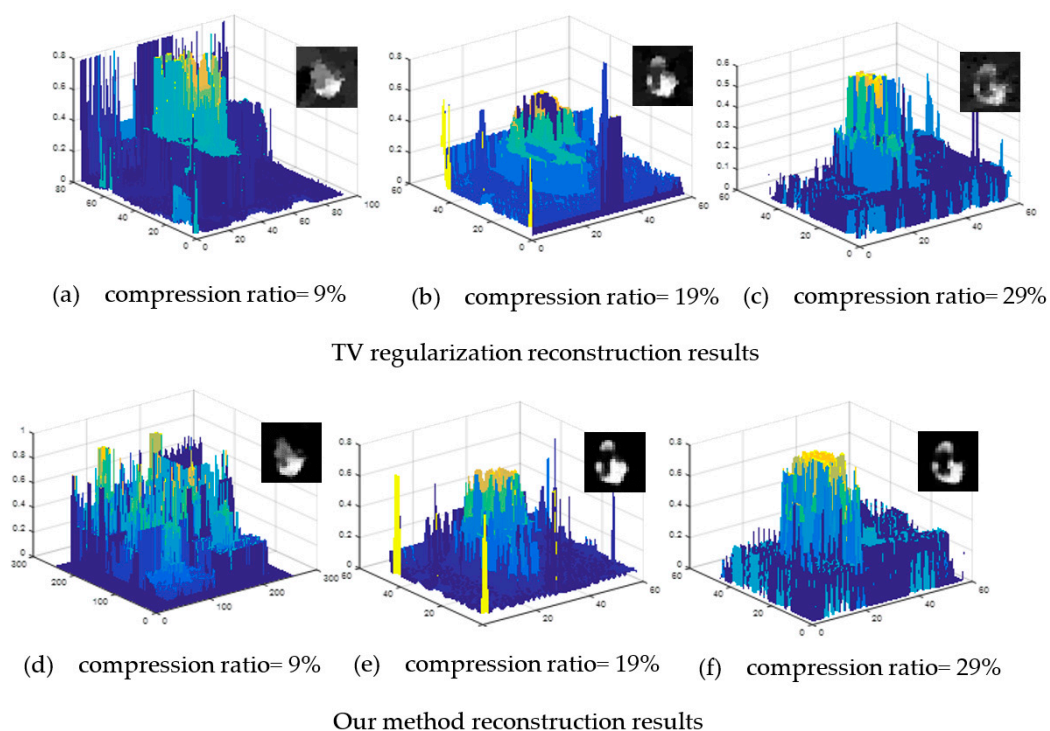


Figure 13. Reconstruction algorithm comparisons: (a–c) TV regularization reconstruction results at different compression ratios and (d–f) our method reconstruction results at different compression ratios.

5. Conclusions

In the current study, an improved underwater SPI system and an image super-resolution method are proposed, which are cost-effective for underwater imaging. This system can overcome the limitations of obtaining good-quality images in low-light conditions. The proposed SPI system employed a combination of the CS algorithm and a neural network to enhance the quality of the images to reduce sampling time and post-processing of data. Referring to the network structure of an CS-SRCNN, the employed network algorithm successfully deduced the HR images from being trained with multiple measurement data. The resultant 2D underwater image using the proposed single-pixel imaging setup and network algorithm ensures better quality compared to the conventional SRCNN method. The experimental results show that, while the PSNR and SSIM of SPI images reach 11.57 dB and 0.26, respectively, in clear water, the PSNR and SSIM of our method image reach 15.67 dB and 0.45. Therefore, the application of network algorithms in our work improves the accuracy of the overall SPI system. Although our method can improve the image quality to some extent, there are limitations, mainly, blur in the resultant image, which will be addressed in the future. Also, we will investigate the approaches suitable for turbid water and will expand the structure of the CS-SRCNN to improve image quality further.

Author Contributions: All authors contributed to the article. M.L. proposed the concept, and conceived and designed the optical system under the supervision of X.W., M.L., J.W.Y., S.L.H.L., and A.M., who performed the experiments, analyzed the data, and wrote the paper. X.W., and X.X. reviewed the manuscript and provided valuable suggestions. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by Ministry of Higher Education, Malaysia under grant FRGS/1/2020/ICT02/MUSM/02/1 and in part by the National Natural Science Foundation of China (No.61903048).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: We are very grateful to the support from School of Engineering, Monash University Malaysia and Changchun university of science and technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lu, H.; Li, Y.; Uemura, T.; Kim, H.; Serikawa, S. Low illumination underwater light field images reconstruction using deep convolutional neural networks. *Future Gener. Comput. Syst.* **2018**, *82*, 142–148. [[CrossRef](#)]
2. Zhang, Y.; Li, W.; Wu, H.; Chen, Y.; Su, X.; Xiao, Y.; Wang, Z.; Gu, Y. High-visibility underwater ghost imaging in low illumination. *Opt. Commun.* **2019**, *441*, 45–48. [[CrossRef](#)]
3. Amer, K.O.; Elbouz, M.; Alfalou, A.; Brosseau, C.; Hajjami, J. Enhancing underwater optical imaging by using a low-pass polarization filter. *Opt. Express* **2019**, *27*, 621–643. [[CrossRef](#)] [[PubMed](#)]
4. Liu, F.; Wei, Y.; Han, P.; Yang, K.; Bai, L.; Shao, X. Polarization-based exploration for clear underwater vision in natural illumination. *Opt. Express* **2019**, *27*, 3629–3641. [[CrossRef](#)]
5. Mariani, P.; Quincoces, I.; Haugholt, K.; Chardard, Y.; Visser, A.; Yates, C.; Piccinno, G.; Reali, G.; Risholm, P.; Thielemann, J. Range-gated imaging system for underwater monitoring in ocean environment. *Sustainability* **2019**, *11*, 162. [[CrossRef](#)]
6. Chang, A.; Jung, J.; Um, D.; Yeom, J.; Hanselmann, F. Cost-effective Framework for Rapid Underwater Mapping with Digital Camera and Color Correction Method. *KSCE J. Civ. Eng.* **2019**, *23*, 1776–1785. [[CrossRef](#)]
7. Lu, J.; Li, N.; Zhang, S.; Yu, Z.; Zheng, H.; Zheng, B. Multi-scale adversarial network for underwater image restoration. *Opt. Laser Technol.* **2019**, *110*, 105–113. [[CrossRef](#)]
8. Tang, C.; Von Lukas, U.F.; Vahl, M.; Wang, S.; Wang, Y.; Tan, M. Efficient underwater image and video enhancement based on Retinex. *Signal, Image Video Process.* **2019**, *13*, 1011–1018. [[CrossRef](#)]
9. Çelebi, A.T.; Ertürk, S. Visual enhancement of underwater images using empirical mode decomposition. *Expert Syst. Appl.* **2012**, *39*, 800–805.

10. Li, M.; Mathai, A.; Yandi, L.; Chen, Q.; Wang, X.; Xu, X. A brief review on 2D and 3D image reconstruction using single-pixel imaging. *Laser Phys.* **2020**, *30*, 095204. [[CrossRef](#)]
11. Mathai, A.; Guo, N.; Liu, D.; Wang, X. 3D Transparent Object Detection and Reconstruction Based on Passive Mode Single-Pixel Imaging. *Sensors* **2020**, *20*, 4211. [[CrossRef](#)] [[PubMed](#)]
12. Mathai, A.; Wang, X.; Chua, S.Y. Transparent Object Detection Using Single-Pixel Imaging and Compressive Sensing. In Proceedings of the 2019 13th International Conference on Sensing Technology (ICST), Sydney, Australia, 2–4 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
13. Li, M.; Mathai, A.; Xu, X.; Wang, X. Non-line-of-sight object detection based on the orthogonal matching pursuit compressive sensing reconstruction. In *Optics Frontier Online 2020: Optics Imaging and Display*; International Society for Optics and Photonics: Bellingham, WA, USA, 2020; p. 115710G.
14. Ouyang, B.; Dalgleish, F.R.; Caimi, F.M.; Giddings, T.E.; Shirron, J.J.; Vuorenkoski, A.K.; Nootz, G.; Britton, W.; Ramos, B. Underwater laser serial imaging using compressive sensing and digital mirror device. In *Laser Radar Technology and Applications XVI*; International Society for Optics and Photonics: Bellingham, WA, USA, 2011; p. 803707.
15. Chen, Q.; Mathai, A.; Xu, X.; Wang, X. A study into the effects of factors influencing an underwater, single-pixel imaging system's performance. In *Photonics*; Multidisciplinary Digital Publishing Institute: Bazel, Switzerland, 2019; p. 123.
16. Chen, Q.; Chamoli, S.K.; Yin, P.; Wang, X.; Xu, X. Active Mode Single Pixel Imaging in the Highly Turbid Water Environment Using Compressive Sensing. *IEEE Access* **2019**, *7*, 159390–159401. [[CrossRef](#)]
17. Chen, Q.; Yam, J.W.; Chua, S.Y.; Guo, N.; Wang, X. Characterizing the performance impacts of target surface on underwater pulse laser ranging system. *J. Quant. Spectrosc. Radiat. Transf.* **2020**, *255*, 107267. [[CrossRef](#)]
18. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [[CrossRef](#)] [[PubMed](#)]
19. Hassannejad, H.; Matrella, G.; Ciampolini, P.; De Munari, I.; Mordonini, M.; Cagnoni, S. Food image recognition using very deep convolutional networks. In Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, Amsterdam, The Netherlands, 16 October 2016; ACM: New York, NY, USA, 2016; pp. 41–49.
20. Caramazza, P.; Boccolini, A.; Buschek, D.; Hullin, M.; Higham, C.F.; Henderson, R.; Murray-Smith, R.; Faccio, D. Neural network identification of people hidden from view with a single-pixel, single-photon detector. *Sci. Rep.* **2018**, *8*, 11945. [[CrossRef](#)]
21. Han, J.; Zhang, D.; Cheng, G.; Liu, N.; Xu, D. Advanced deep-learning techniques for salient and category-specific object detection: A survey. *IEEE Signal. Process. Mag.* **2018**, *35*, 84–100. [[CrossRef](#)]
22. Guan, C.-Z. Realtime Multi-Person 2D Pose Estimation using ShuffleNet. In Proceedings of the 2019 14th International Conference on Computer Science & Education (ICCSE), Toronto, ON, Canada, 19–21 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 17–21.
23. Higham, C.F.; Murray-Smith, R.; Padgett, M.J.; Edgar, M.P. Deep learning for real-time single-pixel video. *Sci. Rep.* **2018**, *8*, 2369. [[CrossRef](#)]
24. Rizvi, S.; Cao, J.; Zhang, K.; Hao, Q. Improving Imaging Quality of Real-time Fourier Single-pixel Imaging via Deep Learning. *Sensors* **2019**, *19*, 4190. [[CrossRef](#)]
25. Bi, Y.; Xu, X.; Chua, S.Y.; Chow, E.M.T.; Wang, X. Underwater Turbulence Detection Using Gated Wavefront Sensing Technique. *Sensors* **2018**, *18*, 798. [[CrossRef](#)]
26. Dutta, R.; Manzanera, S.; Gambín-Regadera, A.; Irlles, E.; Tajahuerce, E.; Lancis, J.; Artal, P. Single-pixel imaging of the retina through scattering media. *Biomed. Opt. Express* **2019**, *10*, 4159–4167. [[CrossRef](#)]
27. Jauregui-Sánchez, Y.; Clemente, P.; Lancis, J.; Tajahuerce, E. Single-pixel imaging with Fourier filtering: Application to vision through scattering media. *Opt. Lett.* **2019**, *44*, 679–682. [[CrossRef](#)] [[PubMed](#)]
28. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
29. Li, C. An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing. Ph.D. Thesis, Rice University, Houston, TX, USA, 2010.
30. Yu, W.-K.; Yao, X.-R.; Liu, X.-F.; Li, L.-Z.; Zhai, G.-J. Three-dimensional single-pixel compressive reflectivity imaging based on complementary modulation. *Appl. Optics.* **2015**, *54*, 363–367. [[CrossRef](#)]
31. Candès, E.J. Compressive sampling. In Proceedings of the International Congress of Mathematicians, 2006, Madrid, Spain, 22–30 August 2006; pp. 1433–1452.
32. Umehara, K.; Ota, J.; Ishida, T. Super-resolution imaging of mammograms based on the super-resolution convolutional neural network. *Open J. Med. Imaging* **2017**, *7*, 180. [[CrossRef](#)]
33. Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.-H.; Liao, Q. Deep learning for single image super-resolution: A brief review. *IEEE Trans. Multimed.* **2019**, *21*, 3106–3121. [[CrossRef](#)]
34. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [[CrossRef](#)]
35. Luo, Z.; Yurt, A.; Stahl, R.; Lambrechts, A.; Reumers, V.; Braeken, D.; Lagae, L. Pixel super-resolution for lens-free holographic microscopy using deep learning neural networks. *Opt. Express* **2019**, *27*, 13581–13595. [[CrossRef](#)]
36. Le, M.; Wang, G.; Zheng, H.; Liu, J.; Zhou, Y.; Xu, Z. Underwater computational ghost imaging. *Opt. Express* **2017**, *25*, 22859–22868. [[CrossRef](#)]

-
37. Li, F.; Zhao, M.; Tian, Z.; Willomitzer, F.; Cossairt, O. Compressive ghost imaging through scattering media with deep learning. *Opt. Express* **2020**, *28*, 17395–17408. [[CrossRef](#)]
 38. Ward, C.M.; Harguess, J.; Crabb, B.; Parameswaran, S. Image quality assessment for determining efficacy and limitations of Super-Resolution Convolutional Neural Network (SRCNN). In *Applications of Digital Image Processing XL*; International Society for Optics and Photonics: Bellingham, WA, USA, 2017; p. 1039605.