

# A joint framework for underwater sequence images stitching based on deep neural network convolutional neural network

Mingwei Sheng<sup>1</sup>, Songqi Tang, Zhuang Cui, Wanqi Wu and Lei Wan

## Abstract

Panoramic stitching technology provides an effective solution for expanding visual detection range of the autonomous underwater vehicle. However, absorption and scattering of light in the water seriously deteriorate the underwater imaging in terms of distance and quality, especially the scattering sharply decreases the underwater image contrast and results in serious blur. This reduces the number of matching feature points between the underwater images to be stitched, while fewer matched points generated make image registration and stitching difficult. To solve the problem, a joint framework is established, which firstly involves a convolutional neural network-like algorithm composed of a symmetric convolution and deconvolution framework for underwater image enhancement. Then, it proposes an improved convolutional neural network-random sample consensus method based on VGGNet-16 framework to generate more correct matching feature points for image registration. The fusion method based on Laplacian pyramid is applied to eliminate artificial stitching traces and correct the position of stitching seam. Experimental results indicate that the proposed framework can restore the color and detail information of underwater images and generate more effective and sufficient matching feature points for underwater sequence images stitching.

## Keywords

Image stitching, AUV, underwater image, convolutional neural network, image enhancement, image registration

Date received: 25 January 2020; accepted: 17 February 2020

Topic: Vision Systems

Topic Editor: Yangquan Chen

Associate Editor: Xiao Liang

## Introduction

The marine resources development and underwater exploration have become an important development strategy for all countries. However, poor visibility, complex terrain environment, and high water pressure limit underwater operation.<sup>1</sup> Underwater operation needs to be completed with the help of professional underwater equipment, and thus underwater vehicles are emerged at the right moment. Underwater detection operation can be implemented for decoupling operation and long-term and large-scale autonomous underwater navigation, detection, and collision

avoidance, which has the advantages of high efficiency, strong controllability, safety, and intelligence.<sup>2–5</sup> When underwater exploration is carried out with autonomous

Science and Technology on Underwater Vehicle Laboratory, Harbin Engineering University, Harbin, China

### Corresponding author:

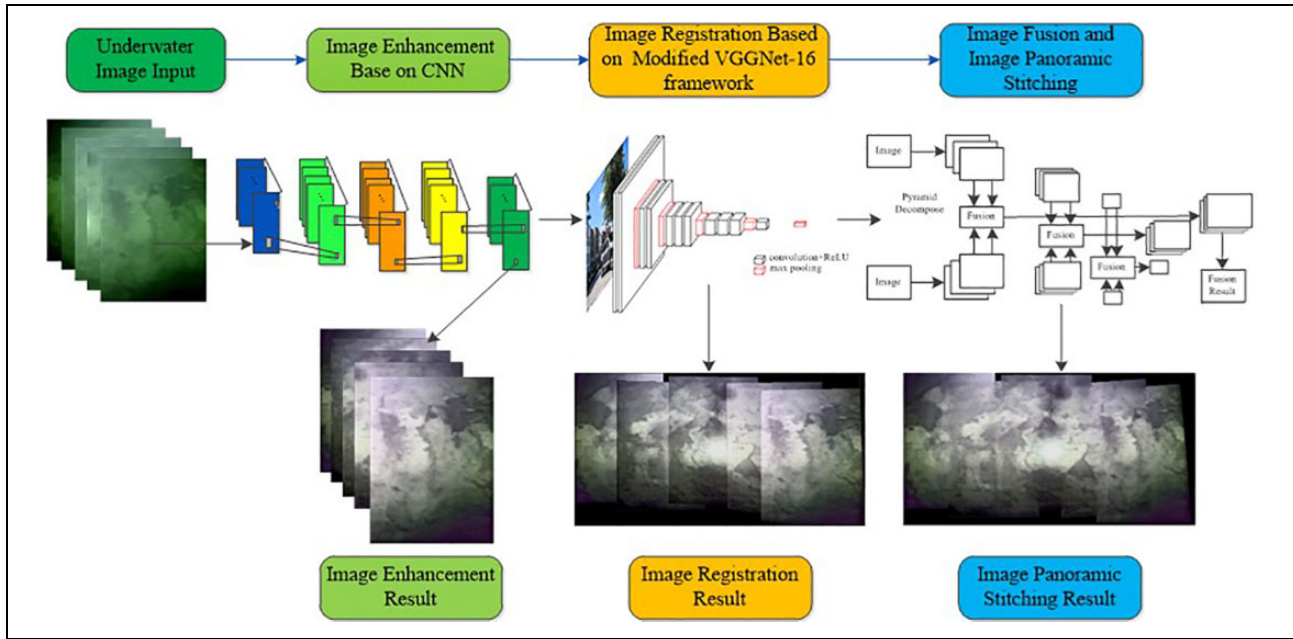
Mingwei Sheng, Science and Technology on Underwater Vehicle Laboratory, Harbin Engineering University, Harbin 150001, China.

Email: smwsky@hrbeu.edu.cn



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without

further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).



**Figure 1.** CNN-based underwater image enhancement and stitching framework. CNN: convolutional neural network.

underwater vehicle (AUV) well environmental perception is the prerequisite for underwater exploration.<sup>6,7</sup> AUV has full application for the fields such as drawing of seabed topography,<sup>8</sup> detection of submarine pipelines,<sup>9</sup> exploration of seabed mineral resources,<sup>10</sup> and visual navigation of underwater vehicle.<sup>11</sup>

Underwater vision is crucial for AUV to obtain environmental information. High underwater imaging quality is conducive to underwater operation.<sup>12</sup> However, the complex underwater environment and poor imaging conditions seriously degrade the image quality obtained by AUV underwater camera, including color attenuation, noise, blur, and low contrast. The attenuation and scattering of light in water also cause color distortion and make the image appear blue-green.<sup>13</sup> The forward and backward scattering of light in underwater transmission limits the contrast and saturation of image.<sup>14</sup> Additionally, the auxiliary lighting on AUV makes the brightness of underwater image uneven, and the impurities, such as organic matter and suspended particles in water, also reduce the quality of underwater image.<sup>15</sup> Since the limited field of view of single underwater image, it is difficult to obtain sufficient underwater information,<sup>16</sup> which greatly limits the visible range and resolution ratio of the underwater optical vision images, thus, the underwater information obtained is insufficient.

In recent years, the research of underwater image enhancement and panoramic image stitching has made great progress, but the complex underwater environment is still a huge challenge. Additionally, there are many difficulties in scientific research to be solved, such as the color fading underwater image still blur after restoration, hard to restore the image with uneven illumination, lack of

effective feature information for panoramic image stitching for seriously degraded underwater image,<sup>17</sup> fewer feature points and high matching error rate leading to the difficulty of underwater image stitching. However, convolutional neural network (CNN) has developed rapidly, and its powerful nonlinear mapping ability has made a breakthrough in computer vision.<sup>18</sup> In this article, CNN is applied to image enhancement and stitching. Firstly, CNN is used to enhance the blurred and degraded underwater image to get a clear image; then, extract and match the feature points of the clear underwater image by VGGNet. With the help of image enhancement, registration and fusion methods, several underwater images are quickly stitched to get a clear panoramic image with wide field of view and high quality for underwater target detection and maintenance.

The proposed joint framework for underwater sequence images stitching based on CNN is shown in Figure 1. In the joint framework, underwater sequences images with the effective overlapping area are selected as input, and the CNN-based enhancement method is firstly performed for effectively improving the quality of underwater images. Based on the improved VGGNet-16, feature extraction is performed in the matching images to generate robust multiple scale feature descriptors and feature points. And then rough feature matching and dynamic interior point selection algorithm is implemented in the two feature sets to generate a rough set of matching feature points pair. An improved RANSAC algorithm is involved to eliminate the mismatched points, and an initial image registration is performed with the calculated homography matrix. Finally, a Laplacian pyramid fusion algorithm is utilized for fusion processing to eliminate the stitching trace for further panoramic stitching of multiple underwater image sequences.

The remainder of this article is organized as follows. The second section describes related works. In the third section, we propose the joint framework including the underwater enhancement, registration based on deep neural network CNN, and a fusion method for stitching. In the fourth section describes the experimental results and discussions. Our conclusions are presented in the fifth section.

## Related work

With the development of computer vision and image processing technology, the purpose of underwater image enhancement is to improve underwater image quality by solving the common phenomena of underwater image, such as color fading, low contrast, and blurred details, so as to improve the accuracy of subsequent image processing. The processing methods for enhancement or restoration of underwater image can be divided into two parts: spatial domain enhancement and frequency domain enhancement.<sup>19</sup>

To improve the visual effect and restore the detail of underwater image, Wang et al.<sup>20</sup> proposed an improved multidimension Retinex color restoration algorithm. The algorithm provides dynamic range compression, local brightness and contrast enhancement, and better color reproduction simultaneously, and further improves the contrast of image. Voronin et al.<sup>21</sup> proposed a method combining local image processing and global image processing in frequency domain. The method applies the logarithmic histogram and spatial equalization method to different image blocks, and the obtained image is the weighted average value of all processing blocks driven by the optimized enhancement measure (EME). Chiang and Chen<sup>22</sup> proposed an underwater image enhancement method based on wavelength compensation and dark channel prior defogging, which applied wavelength compensation to the dark channel prior defogging model for the first time, and the processed image performed well. However, making use of single wavelength in the calculation easily results in color distortion.

The purpose of image stitching is to get an image with larger field of view, higher quality, and better resolution ratio, and to include full details of previous images.<sup>23</sup> At present, the research trend is to improve the quality, robustness, and image stitching speed. Chen et al.<sup>24</sup> proposed an underwater image stitching algorithm based on scale invariant feature transform (SIFT) and wavelet fusion, they considered the poor visibility, imbalanced illumination and viewpoint change are the main factors affecting image feature matching, and then they made full use of the wavelet fusion for obtaining good robustness and accuracy of image feature matching. Babu and Santha<sup>25</sup> proposed an automatic registration and stitching method of deep-sea image based on the replacement features. Harris algorithm is used to extract the feature points in the reference image and detect the image. Apply biorthogonal multiwavelet transform to process the disparity of feature vectors. Then, the

transformation factors are obtained by processing the least square rules supported by the general wavelet transformation. As a result, the image is resampled and transformed to achieve image registration and image stitching.

Lack of data is the biggest challenge of CNN in underwater image processing. Researchers have applied CNN to underwater image enhancement and image stitching since 2017. Wang et al.<sup>26</sup> proposed CNN-based end-to-end framework for underwater image enhancement. The color correction network outputs the color absorption coefficients in different channels to correct the color distortion of underwater image. The defogging network outputs the map of light attenuation transmission to enhance the contrast of underwater image and adopts the pixel interference strategy to effectively improve the convergence speed and accuracy. However, it needs to be realized in way of block overlapping and the calculation cost is higher. Lu et al.<sup>27</sup> proposed using depth convolution neural network to estimate the depth of underwater image, so as to solve the scattering problem in the light field, and to achieve the light field restoration of low-illumination underwater image, but the enhancement effect for some specific scenes is poor. Fabbri et al.<sup>28</sup> proposed untraceable generative adversarial networks (UGAN) network, which makes use of the generative adversarial networks (GAN) to improve the visual quality of underwater image. Using the low-quality underwater data set generated by CycleGAN to train UGAN can effectively solve the problem of color distortion, but the detail information is difficult to recover. Ye et al.<sup>29</sup> proposed an image registration method based on CNN features and SIFT. Fine-tuned the VGG16 model pertained on ImageNet by using a custom data set, so as to obtain CNN features, and to construct the features combined with CNN features and SIFT. Then the combined features were integrated into particle sieve algorithm for image registration, but it was not suitable for underwater image stitching with fewer feature points.

However, the difficulty of applying CNN to underwater image enhancement is the lack of enough underwater image data,<sup>27,30</sup> and the current CNN framework is applicable to few underwater scenes. Furthermore, complex underwater imaging environment and lighting conditions increase the difficulty of underwater image enhancement. Therefore, it is urgent to develop the data set and CNN framework for underwater image enhancement.

Apply the extraction and representation of the CNN model pertained by large-scale data set (ImageNet) to overcome the instability of low-level features and to improve the reliability of registration. Different convolution layers in CNN have different feature description capabilities, which can extract features from different depth layers. The weight sharing of convolution operation effectively reduces the amount of training and provides the possibility of parallelization for training. VGGNet proves that the network performance can be improved by increasing the depth of the network. Use  $3 \times 3$  convolution kernel and  $2 \times 2$



**Figure 2.** Some images of underwater data set.

pooling kernel. The smaller convolution kernel deepens network structure and makes the training results more discriminative through a large number of diverse image data training. It has simple structure and only constructs by superimposing convolution layer, pooling layer and fully connected layer, without branch or quick connection. The structure enables the network to be used for different purposes, including image feature extraction.

## Underwater sequence images processing and stitching method

### Underwater image enhancement based on CNN

*Establishment of the underwater image data set.* The number of underwater images is relatively limited for deep learning, so it is difficult to establish an underwater data set, and collect water degraded images and waterless clear images at the same location are not easy. For deep learning, a large number of training data set is the key factor for parameter training, which is also a bottleneck factor to limit the application of CNN in underwater image. In order to solve this problem, we carry out blur processing on the clear image to generate simulated underwater image. The transmittance  $t$  in the dark channel is evaluated firstly, and then the clear image  $J$  and transmittance  $t$  is utilized, and background light coefficient  $A$  is randomly set in a range  $[0.85, 1]$  to generate a blurred image  $I$

$$I(x) = t(x)J(x) + A(1 - t(x)) \quad (1)$$

The color of the blurred image  $I_{\text{water}}^c$  is attenuated adopting the following formula

$$I_{\text{water}}^c = \alpha^c * I_{\text{blur}}^c \quad (2)$$

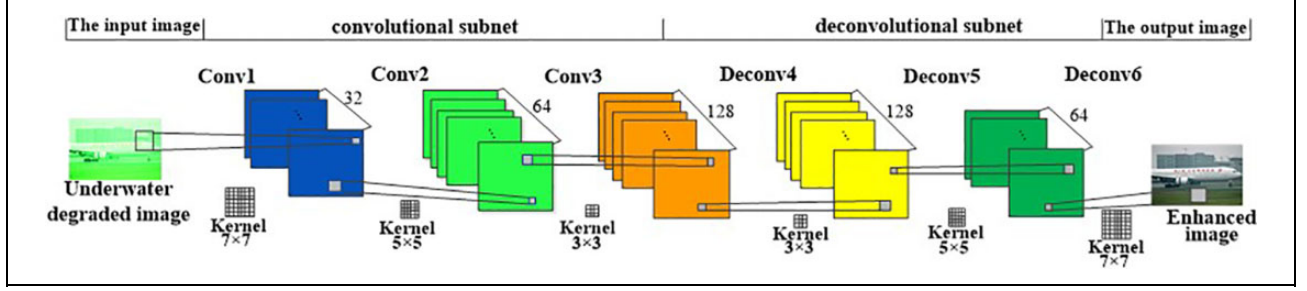
where  $\alpha^c$  represents the attenuation coefficient of different channels, “\*” represents the convolution operation,  $I_{\text{blur}}^c$  represents the blurred image, and  $c$  represents the image channel. By changing the attenuation coefficients in red,

blue, and green channels of the blurred image, it can generate different simulated underwater images with different attenuation degree.

In order to enlarge the variety of underwater images, CycleGAN is also involved to synthesize underwater images. CycleGAN model is trained in an unsupervised way adopting a batch of images from the source domain and the target domain without association. Some underwater images in our data set are shown in Figure 2. All images in the first row are the original images, and the images in the second row are simulated degraded underwater images. The first and two images from left to right in the second row are simulated images generated using blur processing and color attenuation, and the last three images are simulated images synthesized utilizing CycleGAN model.

*CNN-based network framework for underwater image enhancement.* A symmetric convolution and deconvolution CNN-based framework for image enhancement is presented in Figure 3. We continuously adjust the size of convolution kernels and the number of feature maps for optimizing the ability of learning degraded features of the neural network to obtain the nonlinear mapping from underwater degraded image to enhanced image. The blur degraded underwater image is taken as input and the size of the image is unlimited. The hidden layer of presented network is made up of feature maps, including three convolution subnets and deconvolution subnets with different convolution kernels, and the two subnets are symmetrical. The convolution layer is performed to extract features, and the fuzzy degraded features can be learned from the underwater degraded images. Multiple feature maps are generated from convolution kernels. Due to the continuous convolution of CNN fails to restore the details of low-quality underwater images, the symmetrical deconvolution layers are involved to refine the extracted texture features, and it can reconstruct the original image using the feature





**Figure 3.** Proposed CNN-based framework for underwater enhancement. CNN: convolutional neural network.

maps output from convolution layers to generate more details for improving the underwater image quality. After the training by the convolution network, the enhanced image with the same size with the input image is generated in the output layer.

To evaluate the performance of the proposed enhancement method, we propose an evaluation method with reference to mean squared error (MSE). MSE is the mean square value of the difference between the original image and the distorted pixel. Based on the MSE, we calculate Loss, and its formula is as follows

$$\text{Loss} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{wh} \sum_{j=1}^w \sum_{k=1}^h \|D_i(j, k) - X_i(j, k)\|^2 \right) \quad (3)$$

where  $D$  is the enhanced result image,  $X$  is the clear image,  $n$  is the number of training samples,  $W$  and  $H$  are the width and height of training image samples, respectively. In this network, the width and height of training samples remain the same.

During the training process, the loss function is calculated between the trained samples with the corresponding original clear images, and the network parameters are optimized through standard back propagation algorithm and random gradient descent method. The process of updating weights in the network is as follows

$$W_{t+1}^l = W_t^l + \Delta_{t+1}^l \quad (4)$$

$$\Delta_{t+1}^l = -\eta \frac{\partial L}{\partial W_t^l} \quad (5)$$

where  $t$  is the number of iterations,  $l$  is the number of layers,  $\Delta_t^l$  is the weight update value of iteration  $t$  of the  $l$ th layer,  $\eta$  is the learning rate of the  $l$ th layer, and  $\frac{\partial L}{\partial W_t^l}$  is to calculate the partial derivative of the weight in the cost function. The learning rate here is set as  $10^{-4}$ , the weight is initialized with the Gaussian random distribution with the mean value as 0 and the standard deviation as 0.001, and all offsets are initialized as 0.

The fragmented learning approach is involved to generate  $55 \times 55$  pieces of training data from the degraded image and the corresponding clear image by random sampling respectively to improve the efficiency and accuracy

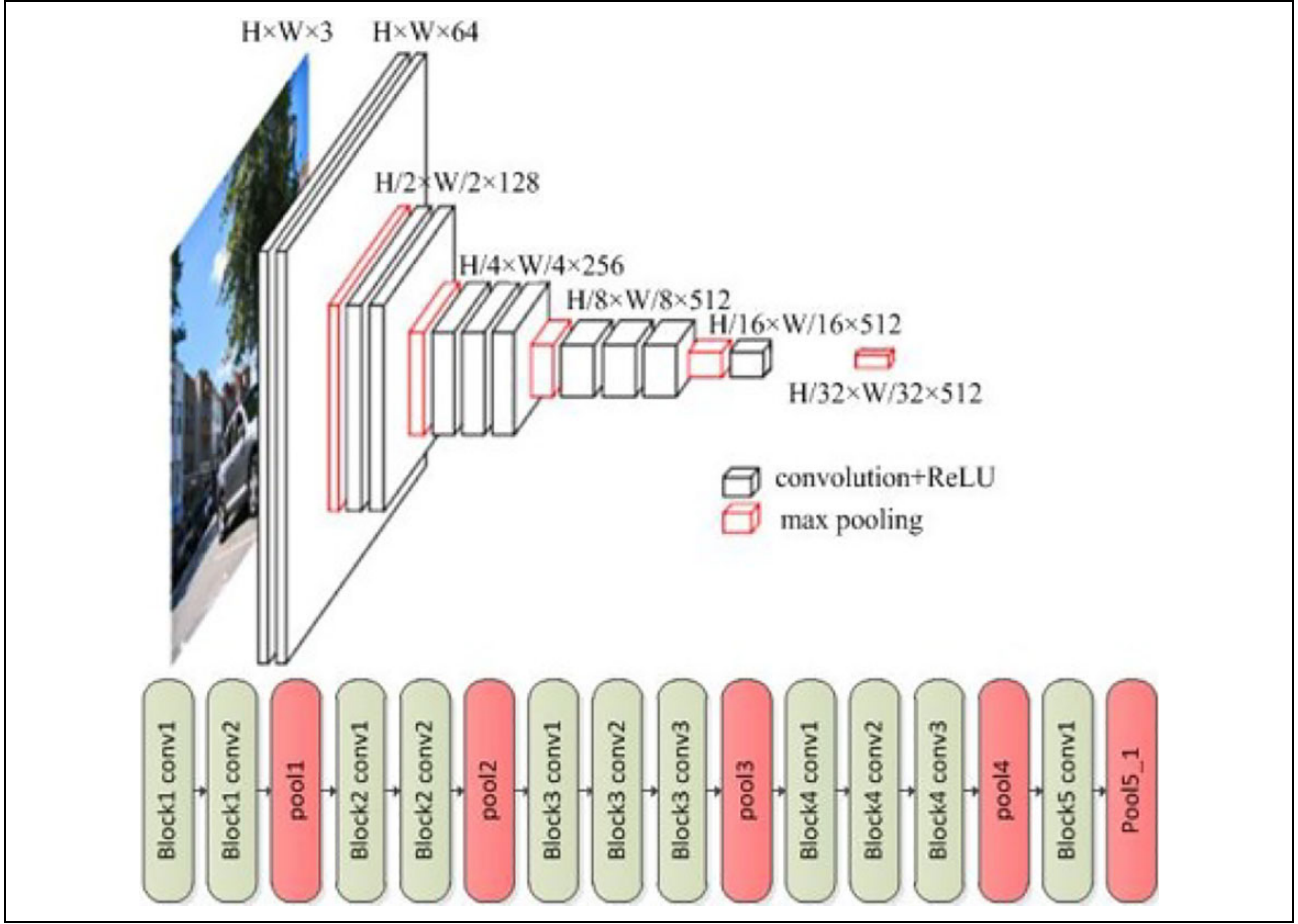
of network training. During the training, the blurred degraded image and the clear image are both input to form the end-to-end mapping for training the proposed CNN.

### Underwater image registration

An underwater image registration method based on improved CNN-RANSAC is proposed. VGGNet-16 framework is trained using the transfer learning in underwater image classification data set, and then more robust multi-scale feature descriptors and feature points are generated by the adjusted VGGNet-16 framework. After rough registration of feature points and the dynamic interior point selection, an improved RANSAC algorithm is utilized to eliminate the mismatch pairs to obtain more accurate underwater image registration results.

**The VGGNet-16 framework training based on transfer learning.** VGGNet-16 model is an image classification network performing classification of 1000 categories. The VGGNet-16 network pretrained by ImageNet is trained by the transfer learning of underwater image classification data set. The parameters of each network are fine-tuned to make it more suitable for underwater image features extraction. The underwater image classification data set includes five categories: sea fish, sea urchin, octopus, coral reef, and jellyfish, and totally includes 1000 pictures. Each image contains only one sample with some color attenuation and fuzzy degradation. The data set is expanded by scaling, translation, flipping, and color dithering to improve the robustness and generalization ability of VGGNet. The learning rate is set to 0.01, the batch size is set to 16, and the number of iterations is set to 200,000. The accuracy of underwater classification verification is up to 90.75%.

**Generating the feature descriptors and points based on improved VGGNet-16 framework.** The modified VGGNet-16 frame structure is shown in Figure 4, the full connection layer and softmax layer are removed, and a maximum pool is added after a convolution in the fifth convolution block. Due to the increase of CNN convolution depth, the expression ability of spatial information gradually decreases. So the input image size is adjusted to  $224 \times 224$ , so as to resize the input as an appropriate size of the receptive field



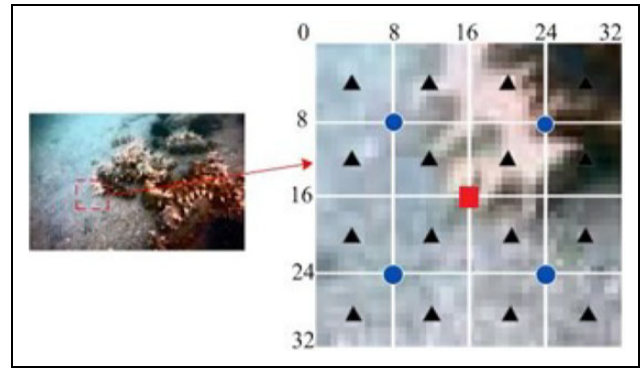
**Figure 4.** Modified VGGNet-16 frame structure.

for reducing the calculation cost. To cover receptive fields with different sizes and generate feature response values, pool3 layer, pool4 layer, and pool5\_1 layer which added after the block5\_conv1 are chose to extract image features.

The output size of the feature map in pool3 layer is  $28 \times 28 \times 256$ , so a grid with a size  $28 \times 28$  is defined to split the whole image block, and each block corresponds to 256 dimensional vectors in the output of the pool3 layer. A feature descriptor is generated from the squares with size of  $8 \times 8$ , and the center of each block is regarded as a feature point. The 256 dimensional vector is defined as the feature descriptor of the pool3 layer, and the output of the pool3 layer is taken as the feature map of the pool3 layer named FM1 with a size of  $28 \times 28 \times 256$ . A pool4 descriptor is generated in each  $16 \times 16$  region of the image, which is shared by four feature points. The pool4 feature map named FM<sub>2</sub> can be calculated using Kronecker product (represented by  $\otimes$ )

$$FM_2 = O_{pool4} \otimes I_{2 \times 2 \times 1} \quad (6)$$

where  $O_{pool4}$  refers to the output of the fourth pooling layer,  $I_{2 \times 2 \times 1}$  presents a tensor with size  $2 \times 2$ . Each descriptor of pool5\_1 is shared by 16 feature points, and the output of



**Figure 5.** CNN-based distribution of feature descriptors and points. CNN: convolutional neural network.

pool5\_1 layer is  $7 \times 7 \times 512$ . The pool5\_1 feature map FM<sub>3</sub> can be calculated as

$$FM_3 = O_{pool5\_1} \otimes I_{4 \times 4 \times 1} \quad (7)$$

Figure 5 shows the distribution of CNN descriptors and feature points of underwater image features in a  $32 \times 32$  rectangle. Each black triangle represents a feature descriptor of pool3 layer, which includes  $8 \times 8$  region. The center

of each region is regarded as a feature point. The blue circle represents the pool4 layer feature descriptor, which includes a  $16 \times 16$  region and can be shared by four feature points. The red square point represents the pool5\_1 layer feature descriptor shared by 16 feature points.

It is necessary to normalize the feature map such as  $FM_1$ ,  $FM_2$ , and  $FM_3$  into unit variance

$$FM_{\text{norm}_i} \leftarrow \frac{FM_i}{\sigma(FM_i)}, i = 1, 2, 3 \quad (8)$$

where  $FM_i$  represents the feature map of the  $i$ th layer,  $\sigma(*)$  is the standard deviation of each value in the calculation matrix,  $FM_{\text{norm}_i}$  represents the characteristic graph normalized to unit variance.

**Rough matching of feature points and selection of dynamic interior points.** The feature distance measure is defined as the feature distance between two feature points  $A$  and  $B$  is the weighted sum of three distance values

$$d(A, B) = \sqrt{2}d_1(A, B) + d_2(A, B) + d_3(A, B) \quad (9)$$

and each component distance value is the Euclidean distance between the respective feature descriptors

$$d_i(A, B) = \text{Euclidean distance}(W_i(A), W_i(B)) \quad (10)$$

where the distance calculated with pool3 descriptors  $d_i(A, B)$  is compensated with a weight  $\sqrt{2}$ , because  $W_1$  is 256 dimensions whereas  $W_2$  and  $W_3$  are 512 dimensions, and  $W_1(A)$ ,  $W_2(A)$ , and  $W_3(A)$  represent the pool3, pool4, and pool5\_1 descriptors of point  $A$ , respectively.

When the following two requirements are met, the feature points  $A$  and  $B$  are considered as a matched points pair:

- (a)  $d(A, B)$  is the smallest of all  $d(*, B)$
- (b) There does not exist a  $d(C, B)$  such that

$d(C, B) \theta \cdot d(A, B)$ , where the matching threshold  $\theta$  is a parameter bigger than 1. The smaller the matching threshold value is, the more feature points are selected.

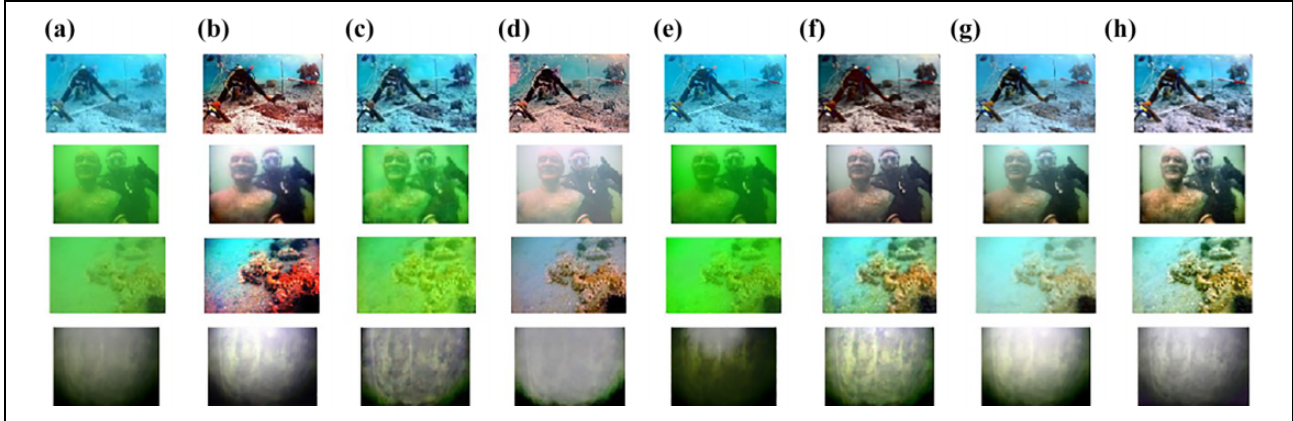
The feature points extracted from the center of the square image block. In the case of deformation, there are partially or completely overlapped image blocks with the corresponding feature points between the reference and registered image. Feature points with higher overlap ratio are considered with a higher matching degree. In order to more accurately match the feature points, the dynamic interior point selection method<sup>31</sup> is involved to determine the degree of alignment. The interior point is reselected in every  $k$  iterations. In the coarse matching stage of feature points, a low threshold value  $\theta_0$  is chose to get a large number of feature points and filter out the irrelevant points. And then specify a large initial threshold value  $\theta'$ , so that only the correct interior point can meet the conditions, and the correct interior point refers to the feature points with overlapping blocks. During the coarse matching process, a threshold  $\theta$  is subtracted from the step size  $\delta$  in every  $k$

iteration, allowing more feature points to affect the transformation. And the feature points of strong matching can determine the overall transformation to improve the accuracy of rough matching of feature points. Here, the threshold value  $\theta_0$  is calculated by 128 pairs of feature point pairs with reliable strong matched points pairs, and  $\theta'$  is calculated by 64 pairs of feature point pairs with the most reliable matched points pairs.

**Improved RANSAC algorithm for elimination of mismatch pairs.** The conventional RANSAC algorithm for elimination of mismatch pairs has two limitations: One is that when there are too many mismatch pairs between the registration images, the number of iterations will greatly increase, which is time-consuming. Another is that in terms of accuracy, the initial model parameter estimation is calculated from the sample data of random sampling, but the selection of the smallest subset is considered from the perspective of efficiency, resulting in nonoptimal model parameters obtained. This article improves the RANSAC algorithm from three aspects as follows:

- (a) Reducing the range of observation data: The matched point pairs are arranged in a database according to the distance from far to near, so the matched point pairs in front of the database have higher registration. We select top 80% matched point pairs after sorting to form a new database, and then calculating the parameter model of homography matrix  $H$ , which will effectively reduce the number of iterations and reduce calculation time.
- (b) Eliminating the cross feature points: In fact, there are no crossing lines between the registered images if we connect corresponding feature points with lines, so if there is a line intersecting with many lines, the point pair is assumed as mismatching. So set a threshold, and a number of intersection between the connecting lines of a matched point pair and other lines can be calculated. If the number surpasses the threshold, the point pair will be eliminating to improve the matching accuracy.
- (c) Removing feature points that are not in the target area: Sometime the feature points in the image target area and background will match, resulting in the increase of mismatch pairs. Therefore, it is necessary to remove the feature points in the background area, and only keep the feature points in the target area. The number of successfully matched points around one point is calculated, if it is less than the threshold value, the point is assumed as a background point, and it will be removed from the interior point set.

After the improved RANSAC algorithm performed, we obtain the left matching feature pairs between the reference



**Figure 6.** Results of different underwater image enhancement methods.

image and registered image and use them to calculate the homography matrix  $H$  adopting the following formula

$$P = HP' = \begin{bmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & 1 \end{bmatrix} P' \quad (11)$$

where  $P$  is expressed as  $(x, y, 1)^T$ ,  $P'$  is expressed as  $(x', y', 1)^T$ ;  $(x, y)$  and  $(x', y')$  are one matched points pair. Here, four random matched point pairs are selected to calculate the homography matrix  $H$ .

### Underwater sequence images stitching method

After image enhancement and registration, images fusion among the registered images is a key step for panoramic underwater sequence images stitching. In this article, a Laplacian pyramid image fusion algorithm is involved. The image is divided into different frequency segments for fusion, the upper layer of the pyramid contains the overall outline of the image, while the lower layers are utilized to analyze the details.

The conventional stitching approach adopts frame by frame splicing method. Through registration of adjacent images, calculation of transformation matrix and multiplication, the transformation matrix between a frame image and panoramic coordinate system is obtained. Panoramic stitching image can be obtained utilizing the transformation matrix to include all images in the panoramic coordinate system. However, this seriously affects the accuracy and quality of panoramic image stitching due to more and more accumulated errors, which also limits the number of images.

## Experimental results and analysis

### Underwater image enhancement results and analysis

In order to verify the effect of the enhancement method in real underwater images, real underwater images with different fuzzy degradation types and scenes are selected for

processing, and the frequently-used methods are selected for comparison, including Histogram Equalization (HE), Contrast Limited Adaptive Histogram Equalization (CLAHE), Multi-Scale Retinex with Color Restore (MSRCR), Dark Channel Prior (DCP), Image Fusion Enhancement (IFE), and Wavelength Compensation and Image Defogging (WCID).

The result of different underwater image enhancement methods is presented in Figure 6, the first two images are typical underwater color attenuation images, and the third image is submarine images taken by AUV, and the fourth image is tunnel wall images captured by AUV for underwater tunnel detection. In terms of the image sharpness improvement, the effect of the proposed method is similar to that of HE, DCP, and MSRCR method, but HE method has partial color distortion and deviates from normal color. The enhancement result of DCP and CLAHE method is still color attenuation, which can be applied to the enhancement of underwater fuzzy image without color attenuation. The image applying MSRCR enhancement method is foggy, which making the details blurred. Although the enhancement results utilizing IFE and WCID methods are closest to the proposed method, which can effectively improve the clarity and color attenuation, some image details are still blurred. Compared with other underwater image enhancement methods, the enhanced results of the proposed method in this article show its superior performance with higher image contrast and clearer details with less noise.

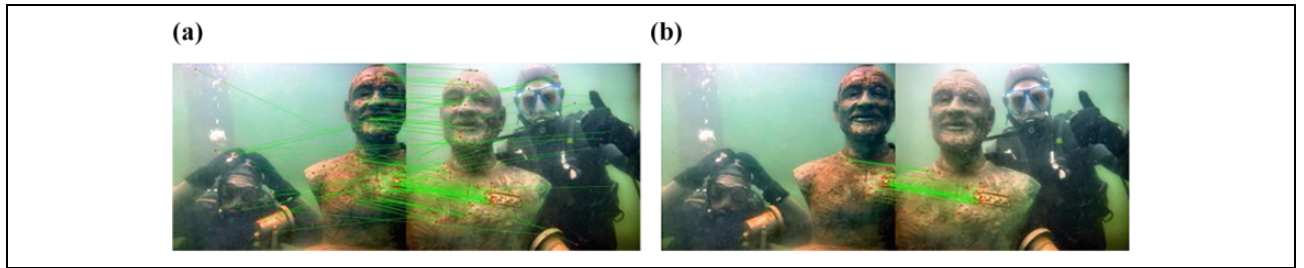
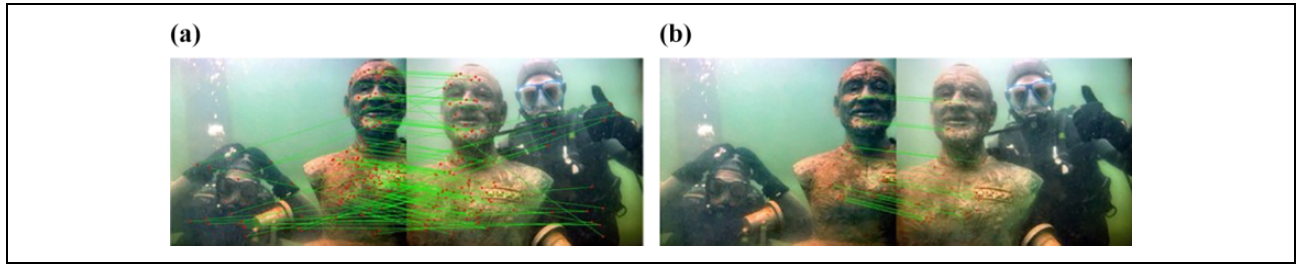
Twenty underwater images are selected for enhancement, and the seven evaluation methods including Mean, Standard, Information Entropy, Blur, Average Gradient, underwater color image quality evaluation (UCIQE), and underwater image quality evaluation metric (UIQM), are all implemented for quantitative calculation. The comparison results of indicators are shown in Table 1, the Mean and Information Entropy of the proposed method are the highest, and Standard result is slightly lower than HE method, and Blur value is the lowest, and Average Gradient, UCIQE, and UIQM value is the highest. Therefore, the proposed method shows higher clarity and contains more



**Table 1.** Comparison of image enhancement evaluation indexes under different methods.

Method	Mean	Standard	Entropy	Blur	Gradient	UCIQE	UIQM
HE	127.346	<b>74.106</b>	7.749	0.248	8.096	0.486	4.267
CLAHE	116.299	44.373	7.409	0.319	4.444	0.321	2.839
MSRCR	129.671	55.804	7.285	0.278	5.897	0.456	3.954
DCP	90.216	66.788	7.341	0.326	3.904	0.357	2.932
UIEF	136.609	60.325	7.806	0.253	7.712	0.435	4.431
WCID	110.700	63.327	7.821	0.223	6.922	0.451	4.597
Proposed method	<b>142.319</b>	68.282	<b>7.859</b>	<b>0.214</b>	<b>8.195</b>	<b>0.495</b>	<b>4.810</b>

HE: Histogram Equalization; CLAHE: Contrast Limited Adaptive Histogram Equalization; MSRCR: Multi-Scale Retinex with Color Restore; DCP: Dark Channel Prior; IFE: Image Fusion Enhancement; WCID: Wavelength Compensation and Image Defogging. Significance for boldface values mean the best value in the category.

**Figure 7.** Coarse and exact matching results based on SIFT method. (a) Coarse matching point pairs. (b) Exact matching point pairs.**Figure 8.** Coarse and exact matching results based on SURF method. (a) Coarse matching point pairs. (b) Exact matching point pairs.

information, which represents better enhancement processing effect.

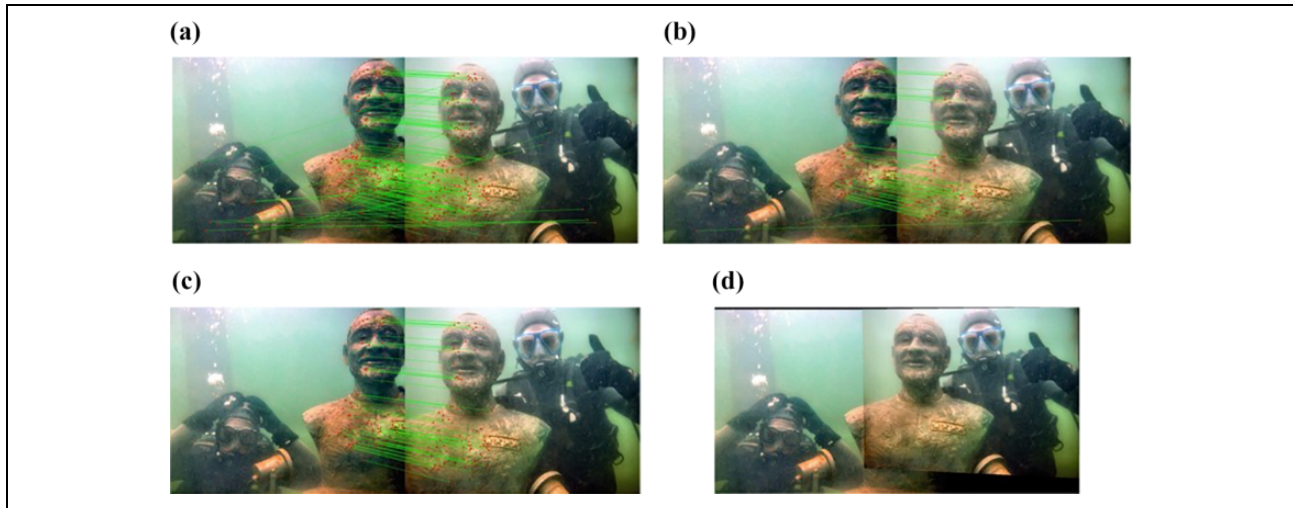
### Underwater image registration results and analysis

The underwater image SIFT-based and speeded-up robust features (SURF)-based registration results are shown in Figures 7 and 8, respectively. The registration results based on the our proposed method are shown in Figure 9. Comparison of data index calculation results of underwater image registration with different registration methods are summarized and shown in Table 2.

As shown in Table 2, in terms of feature extraction, although the SIFT method obtains the largest number of feature points, it consumes more time. The SURF method generates the least feature points, which is about half of the SIFT method, but it has the least time consumption and performs better real-time performance. The feature points extracted by the proposed method in this article are slightly less than SIFT

method, but less time-consuming than SIFT method. Furthermore, when compared with SURF method, more feature points can be extracted, and less time consumed adopts the proposed method. The proposed method in the coarse matching stage increased the dynamic interior point selection, which effectively improves the precision of coarse matching feature points, and generates more coarse matching points pairs. As shown in Figures 7(a) and (b), 8(a) and (b), and 9(a) and (b), the proposed registration method generates fewer mismatch point pairs than SIFT-based and SURF-based registration methods on the rough matching stage and keep more exact matching point pairs on the exact matching stage.

As shown in Figure 9(b) and (c), utilizing the conventional RANSAC method, 54 pairs of precise matching point pairs with 2 pairs of mismatched point pairs are generated, and 76 accurate matching point pairs are obtained utilizing the improved RANSAC method without no errors point pairs. This shows that the improved RANSAC method performs better than the original algorithm.



**Figure 9.** Coarse and exact matching results based on the proposed method. (a) Coarse matching point pairs. (b) Exact matching point pairs. (c) Exact matching point pairs adopting improved RANSAC method. (d) Image registration result based on the proposed method.

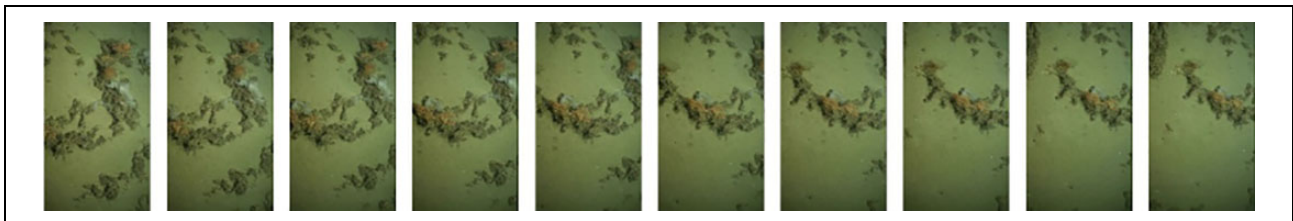
**Table 2.** Comparison of underwater image registration results under different methods.

Method	Image	Features extraction stage		Rough matching stage		Exact matching stage		Interior points ratio (%)
		Number of feature point of	Time(s)	Number of rough matched points pair	Time(s)	Number of exact matched points pair	Time(s)	
SIFT-based	Image A	<b>1039</b>	2.269	129	0.230	45	13.720	34.8
	Image B	<b>912</b>	2.151					
SURF-based	Image A	553	<b>0.757</b>	110	<b>0.158</b>	30	12.671	27.2
	Image B	512	<b>0.726</b>					
Proposed method	Image A	989	1.682	<b>155</b>	0.282	<b>76</b>	<b>10.395</b>	<b>49</b>
	Image B	864	1.634					

Significance for boldface values mean the best value in the category.



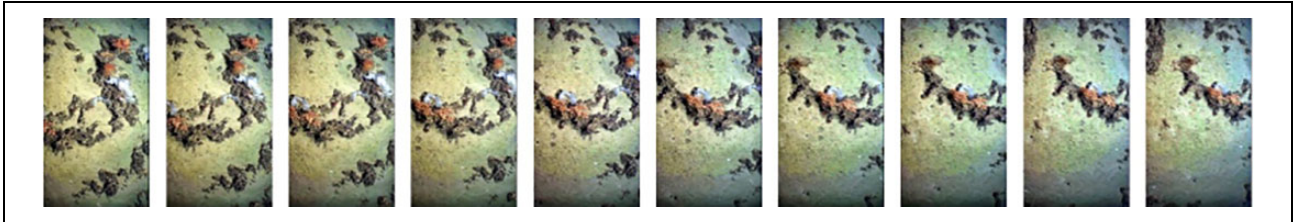
**Figure 10.** The underwater tunnel interior sequence images.



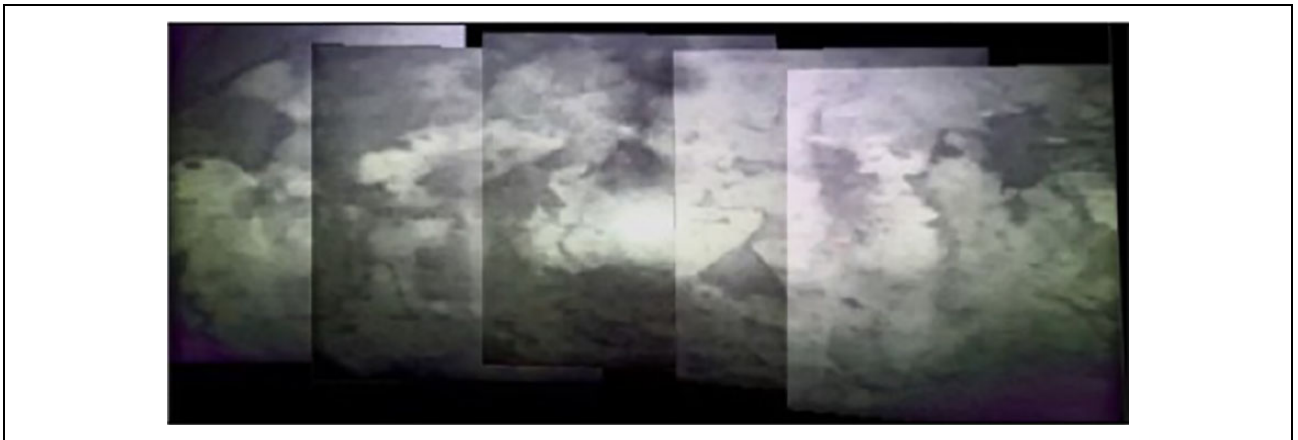
**Figure 11.** The seafloor sequence images.



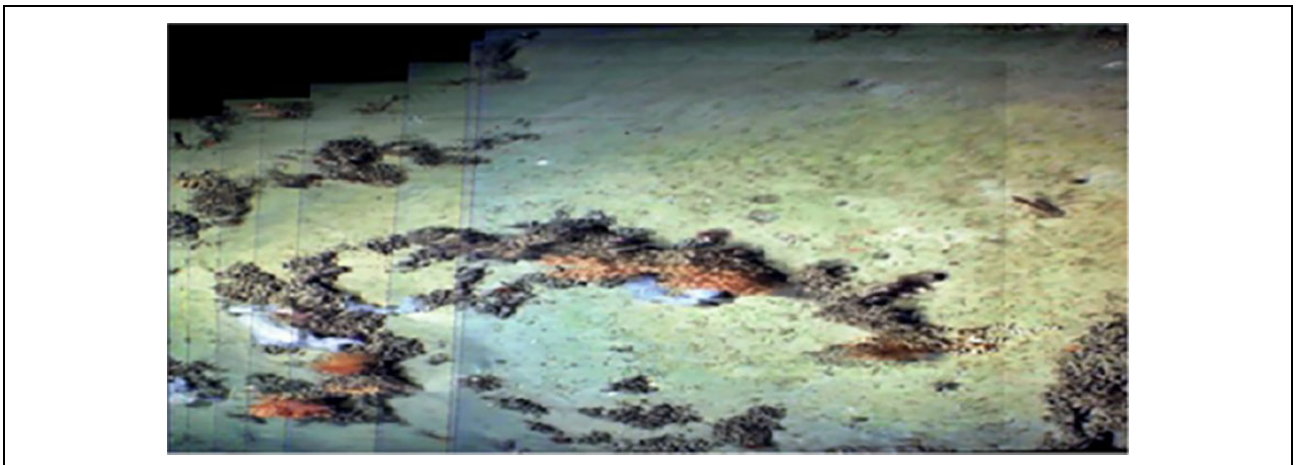
**Figure 12.** The enhanced underwater tunnel interior sequence images.



**Figure 13.** The enhanced sequence seafloor images.



**Figure 14.** The stitching result of underwater tunnel interior images without image fusion.

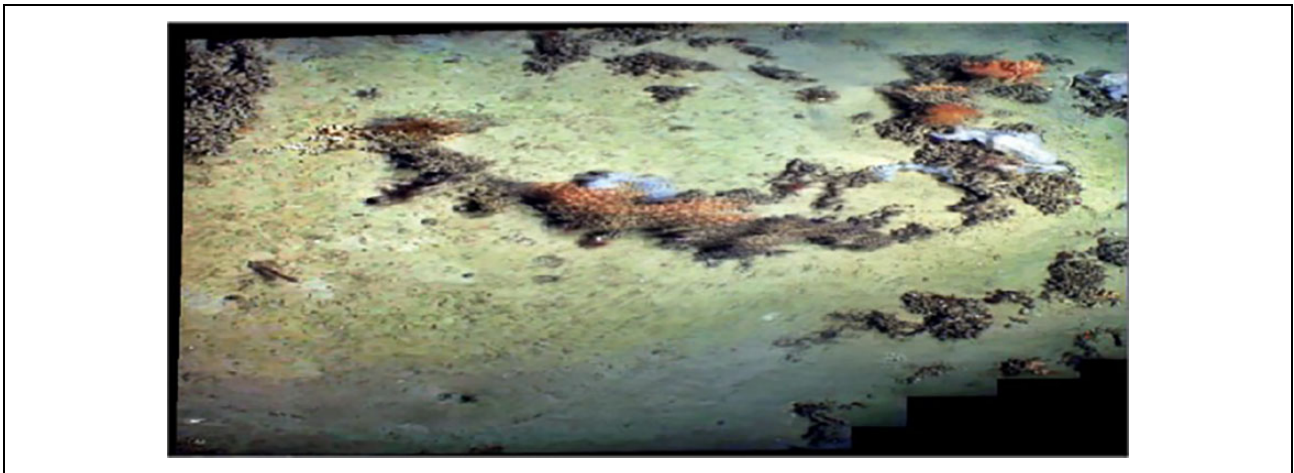


**Figure 15.** The stitching result of seafloor images without image fusion.





**Figure 16.** The underwater tunnel interior panoramic image utilizing image fusion method.



**Figure 17.** The seafloor panoramic image utilizing image fusion method.

### *Underwater panoramic image stitching results and analysis*

Aim to verify the performance of the proposed panoramic image stitching method, some underwater sequence images captured by AUV in some typical underwater environments such as seafloor and underwater tunnel, are selected to generate underwater panoramic images with wide field. Five underwater tunnel interior sequence images are shown in Figure 10, and 10 seafloor sequence images are shown in Figure 11. The proposed CNN-based underwater image enhancement algorithm are implemented on above 15 underwater sequence images, and the enhancement results are shown in Figures 12 and 13, respectively. As shown in Figures 12 and 13, the proposed enhancement method improves the brightness and clarity effectively, which makes the detail information more prominent and obvious, and can recover the blue-green tone caused by the absorption of light. Adopting the proposed CNN-RANSAC registration method and fusion method, the underwater stitching results are shown in Figures 14 and 15, and the panoramic results of underwater tunnel interior panoramic image utilizing image

fusion method are shown in Figures 16 and 17. As shown in Figures 16 and 17, the Laplace pyramid image fusion method effectively eliminate the step change of illumination and the gap between the mosaic images, and make the stitching area continuous transition, and the detail information between the images can be preserved completely.

### **Conclusions**

In this article, we propose a joint framework for panoramic underwater sequence images stitching, that is, an underwater wide range visual perception task, which requires several high quality underwater images and enough matching feature points for registration. For these purposes, we firstly establish an underwater image data set, and construct a CNN framework of symmetric convolution and deconvolution for image enhancement. And then we propose an underwater image registration method based on improved CNN-RANSAC to generate sufficient accurate matching feature points after rough matching of feature points and selection of dynamic interior points. Finally, aiming to eliminate the artificial stitching trace and correct the position of the



stitching seam, a fusion method based on Laplace pyramid is presented and implemented on the enhanced and registered underwater sequence images. The proposed framework is validated utilizing the images captured by underwater camera equipped on AUVs, through different underwater environment experiments including seafloor and pressure water tunnel. The effectively stitching performance of the proposed joint framework is demonstrated.

In future work, the framework for underwater sequence images stitching based on CNN article will be implemented in the embedded system equipped on AUV to perform real-time stitching.


### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported in part by the National Natural Science Foundation of China under Grant (51979057, 51979058, 51609050), in part by the Defense Industrial Technology Development Program (JCKYS2019604SXJQR-09), in part by the Research Fund from Science and Technology on Underwater Vehicle Technology (6142215180209), and in part by the Fundamental Research Funds for the Central Universities Facing International Academic Frontier Support Program (3072019CFG0101).

### ORCID iD

Mingwei Sheng  <https://orcid.org/0000-0003-0734-3738>

### References

1. Ji DX, Li HC, Chen CW, et al. Visual detection and feature recognition of underwater target using a novel model-based method. *Int J Adv Robot Syst* 2018; 15(6): 1–11.
2. Liang X, Qu XR, Wang N, et al. Swarm control with collision avoidance for multiple underactuated surface vehicles. *Ocean Eng* 2019; 191(106516): 1–10.
3. Qin HD, Yu X, Zhu ZB, et al. An expectation-maximization based single-beacon underwater navigation method with unknown ESV. *Neurocomputing* 2020, 378: 295–303.
4. Liang X, Qu XR, Wang N, et al. A novel distributed and self-organized swarm control framework for underactuated unmanned marine vehicles. *IEEE Access* 2019; 7: 112703–112712.
5. Qin HD, Chen H, Sun YC, et al. Distributed finite-time fault-tolerant containment control for multiple ocean bottom flying node systems with error constraints. *Ocean Eng* 2019; 189: 106341.
6. Yang X, Liu ZY, Qiao H, et al. Underwater image matching by incorporating structural constraints. *Int J Adv Robot Syst* 2017; 14(6): 1–10.
7. Jung J, Choi J, Lee Y, et al. AUV localization using depth perception of underwater structures from a monocular camera. In: *OCEANS 2016 MTS/IEEE Monterey*, Monterey, USA, 19–23 September 2016, pp. 1–4.
8. Li J, Lee M, Kim J, et al. Development of P-SURO II hybrid AUV and its experimental study. In: *OCEANS 2013 MTS/IEEE Bergen*, Bergen, Norway, 10–13 June 2013, pp. 1–6.
9. Khan A, Ali S, Meriaudeau F, et al. Visual feedback-based heading control of autonomous underwater vehicle for pipeline corrosion inspection. *Int J Adv Robot Syst* 2017; 14(3): 1–5.
10. Yokota S, Kim K, and Imasato M. Development and sea trial of an autonomous underwater vehicle equipped with a sub-bottom profiler for surveying mineral resources. In: *IEEE/OES autonomous underwater vehicles (AUV)*, Tokyo, Japan, 6–9 November 2016, pp. 81–84.
11. Hover F, Eustice R, Kim A, et al. Advanced perception navigation and planning for autonomous in-water ship hull inspection. *Int J Robot Res* 2012; 31(12): 1445–1464.
12. Liu RS, Fan X, Zhu M, et al. Real-world underwater enhancement: challenges, benchmarks, and solutions. *IEEE Trans Image Process* 2019: 1–12.
13. Galdran A and Alvarez G. Automatic red channel underwater image restoration. *J Visual Commun Image R* 2015; 26: 132–145.
14. Peng Y and Cosman P. Underwater image restoration based on image blurriness and light absorption. *IEEE Trans Image Process* 2017; 26(4): 1579–1594.
15. Panetta K, Gao C, and Agaian S. Human-visual-system-inspired underwater image quality measures. *IEEE J Oceanic Eng* 2016; 41(3): 541–551.
16. Li HY, Luo J, and Huang CJ. An adaptive image-stitching algorithm for an underwater monitoring system. *Int J Adv Robot Syst* 2014; 11: 1–8.
17. Ruan JX, Xie LY, and Ruan YY. Image stitching algorithm based on SURF and wavelet transform. In: *7th international conference on digital home (ICDH)*, Guilin, China, 30 November–1 December 2018, pp. 9–13.
18. Liu S, Liang X, Liu L, et al. Matching-CNN meets KNN: quasi-parametric human parsing. In: *IEEE conference on computer vision and pattern recognition*, Boston, USA, 7–12 June 2015, pp. 1419–1427.
19. Yang JY, Wang XY, Yue HJ, et al. Underwater image enhancement based on structure-texture decomposition. In: *IEEE international conference on image processing (ICIP)*, Beijing, China, 17–20 September 2017, pp. 1207–1211.
20. Wang JB, Lu K, Xue J, et al. Single image dehazing based on the physical model and MSRCR algorithm. *IEEE T Circ Syst Vid Technol* 2018; 28(9): 2190–2199.
21. Voronin V, Semenishchev E, and Tokareva S. Underwater image enhancement algorithm based on logarithmic transform histogram matching with spatial equalization. In: *14th IEEE international conference on signal processing (ICSP)*, Beijing, China, 12–16 August 2018, pp. 434–438.
22. Chiang JY and Chen YC. Underwater image enhancement by wavelength compensation and dehazing. *IEEE Trans Image Process* 2012; 21(4): 1756–1769.

23. Rahul R, Shishir P, Karen P, et al. Adaptive alpha-trimmed correlation based underwater image stitching. In: *IEEE international symposium on technologies for homeland security (HST)*, Waltham, USA, 25–26 April 2017, pp. 1–7.
24. Chen MM, Nian R, He B, et al. Underwater image stitching based on SIFT and wavelet fusion. In: *OCEANS*, Genoa, Italy, 18–21 May 2015, pp. 1–4.
25. Babu M and Santha T. Efficient brightness adaptive deep-sea image stitching using biorthogonal multi-wavelet transform and Harris algorithm. In: *International conference on intelligent computing and control (I2C2)*, Coimbatore, India, 23–24 June 2017, pp. 1–5.
26. Wang Y, Zhang J, Cao Y, et al. A deep CNN method for underwater image enhancement. In: *IEEE international conference on image processing*, Beijing, China, 17–20 September 2017, pp. 1382–1386.
27. Lu H, Li Y, Uemura T, et al. Low illumination underwater light field images reconstruction using deep convolutional neural networks. *Future Gener Comp Syst* 2018; 82: 142–148.
28. Fabbri C, Islam MJ, and Sattar J. Enhancing underwater imagery using generative adversarial networks. In: *IEEE international conference on robotics and automation (ICRA)*, Brisbane, Australia, 21–25 May 2018, pp. 7159–7165.
29. Ye FM, Su YF, Xiao H, et al. Remote sensing image registration using convolutional neural network features. *IEEE Geosci Remote Sens Lett* 2018; 15(2): 232–236.
30. Fu X, Huang J, Ding X, et al. Clearing the skies: a deep network architecture for single-image rain removal. *IEEE Trans Image Process* 2017; 26(6): 2944–2956.
31. Yang ZQ, Dan TT, and Yang Y. Multi-temporal remote sensing image registration using deep conventional features. *IEEE Access* 2018; 6: 38544–38555.