

Article

# Tracking and Localization based on Multi-angle Vision for Underwater Target

Jun Liu <sup>1,2</sup>, Shenghua Gong <sup>1</sup>, Wenxue Guan <sup>1,\*</sup>, Benyuan Li <sup>1</sup>, Haobo Li <sup>1</sup> and Jiaxin Liu <sup>1</sup>

<sup>1</sup> College of Computer Science and Technology, Jilin University, Changchun 130012, China; liujun1509@jlu.edu.cn (J.L.); gongsh18@mails.jlu.edu.cn (S.G.); liby18@mails.jlu.edu.cn (B.L.); lih19@mails.jlu.edu.cn (H.L.); liujiixin17@mails.jlu.edu.cn (J.L.)

<sup>2</sup> College of Electronic Information Engineering, Beihang University, Beijing 100191, China

\* Correspondence: guanwx18@mails.jlu.edu.cn; Tel.: +86-153-3077-1569

Received: 19 October 2020; Accepted: 5 November 2020; Published: 7 November 2020



**Abstract:** With the cost reduction of underwater sensor network nodes and the increasing demand for underwater detection and monitoring, near-land areas, shallow water areas, lakes and rivers have gradually tended to densely arranged sensor nodes. In order to achieve real-time monitoring, most nodes now have visual sensors instead of acoustic sensors to collect and analyze optical images, mainly because cameras might be more advantageous when it comes to dense underwater sensor networks. In this article, image enhancement, saliency detection, calibration and refraction model calculation are performed on the video streams collected by multiple optical cameras to obtain the track of the dynamic target. This study not only innovatively combines the application of AOD-Net's (all-in-one network) image defogging algorithm with underwater image enhancement, but also refers to the BASNet (Boundary-Aware Salient network) network architecture, introducing frame difference results in the input to reduce the interference of static targets. Based on the aforementioned technologies, this paper designs a dynamic target tracking system centered on video stream processing in dense underwater networks. As part of the process, most nodes carried underwater cameras. When the dynamic target could be captured by at least two nodes in the network at the same time, the target position could then be calculated and tracked.

**Keywords:** underwater sensors network; video saliency detection; dynamic target tracking and localization

## 1. Introduction

In recent years, advances in underwater sensor network have facilitated a wide variety of exciting scientific applications [1–3], including automated surveys of underwater environments and underwater object detection [4–6]. Regarding underwater detection and underwater detection networks, most of its applications—including target detection and tracking, underwater communication network distance recognition, underwater image enhancement and restoration—tend to use physical models to calculate the scattering model and target depth. It is very necessary to obtain the true position of the target in the image and the precise distance between the target and the detection node. At present, there is a large number of underwater-based wireless sensor network node positioning algorithms. Considering whether the angle or distance information between nodes is measured in the position calculation process, the positioning algorithm can be divided into two categories: distance-based and distance-free. Furthermore, there are four typical ranging methods: based on signal arrival time (time of arrival, TOA), signal arrival time difference (time difference of arrival, TDOA), based on signal angle (angle of arrival, AOA) and signal received strength (received) signal strength indicator, RSS [7].

Generally, localization schemes aim to achieve large coverage, low communication, high accuracy, low deployment cost and good scalability [8–10].

It can be seen from the above methods that most of the positioning methods in the underwater acoustic network use acoustic communication, optical communication or ultra-short baseline, requiring high network connection and time synchronization between nodes, to obtain only the location and volume information of target objects. However, in the complex and harsh underwater acoustic environment, the sound detection method cannot guarantee the accurate extraction of target information [11–13]. Moreover, most ranging and positioning methods can be disadvantageous due to the heavy hardware equipment carried by the node and expensive detection costs. In addition, in the process of AUV (Autonomous Underwater Vehicle)-assisted positioning, researchers hope to obtain accurate pose information of the robot instead of simple position information. Therefore, some studies have already examined inertial navigation, but such process might cause accumulated errors due to drift [14,15].

With the development and research of visual positioning, as well as the deployment and cost reduction of optical image components, during the gradual construction of underwater sensor networks, underwater cameras have gradually become one of the main sensors to perceive and detect the environment. That said, this article applies a multi-node underwater camera group to detect, track and locate the target object entering the detection network or the underwater robot in the network. The core of this paper is to design an underwater vision network monitoring system using 2–3 cheap cameras to detect the same target at the same time for visual ranging and positioning, completing underwater detection task. In order to ensure the feature extraction of underwater images, the median filtering method is first used to preprocess the acquired underwater images that are then dehazed and enhanced. The following step is to use several existing saliency detection methods to perform target detection on the optical video collected by multi-source nodes and, after that, compare their detection effects. In addition, it is necessary to select a suitable saliency detection network model and improve it to obtain the position information of underwater targets. The model here proposed can not only detect areas with high saliency in the optical video, but also reduce the interference of static but high saliency objects in the screen by changing the network model to obtain accurate position information of the target, and assist the navigation and positioning of the AUV or simplify the target object detection task. Finally, this paper constructs a mathematical model to convert the target pixel coordinates in the image collected by the multi-source node into world coordinates as a way to complete the calculation and detection of the target position information.

## 2. System Design and Process Introduction

The core of this study is to design a dynamic target detection system for a lightweight dense visual sensor node network. The system consists of three major parts: visual network detection, video stream target detection, and target location tracking. Its structure is as Figure 1.

The detection and tracking system consists of dense hydro-acoustic network nodes made with inexpensive visual sensors. When an unknown target breaks into the visible area or a movable detection node in the network moves into the detection area, target detection and tracking are triggered.

The detection node will notify the neighbor nodes within one hop of this node. When receiving feedback information from the neighbor node that also detects the target, it will bind 2–3 nodes at the same time to track and locate the dynamic target in the area jointly. After binding, the process of target detection starts. The activated 2–3 visible nodes rotate the pan-tilt to collect the video of the target, and intercept the video streams of 2–3 nodes that simultaneously detect the object. Next, median filtering and AOD-Net defogging method are applied to preprocess the image with the target. The detailed description of this stage can be seen in Section 3.1.

Afterwards, the enhanced stream data is calculated by the frame difference method and the result is brought into the BASNet network. Time information is incorporated into the network to help the model detect the saliency area while filtering out the static interference items. Further on, this saliency

detection part will be described in detail in Section 3.2. Subsequently, the detection node transmits the pixel data with the target centroid information to the central processing node.

The last step is the localization part, which will perform the camera's underwater calibration and the calculation of the refraction model before the deployment of the visual network node. The sink node will integrate the internal and external parameters of the camera, the refraction model, and the real-time angle to convert between pixel and world coordinates. Finally, it uses the intersection of multiple cameras and the target to calculate the true position of the target in the world coordinates to achieve tracking and localization function. Details of the outcomes of this process will be described in Section 4.

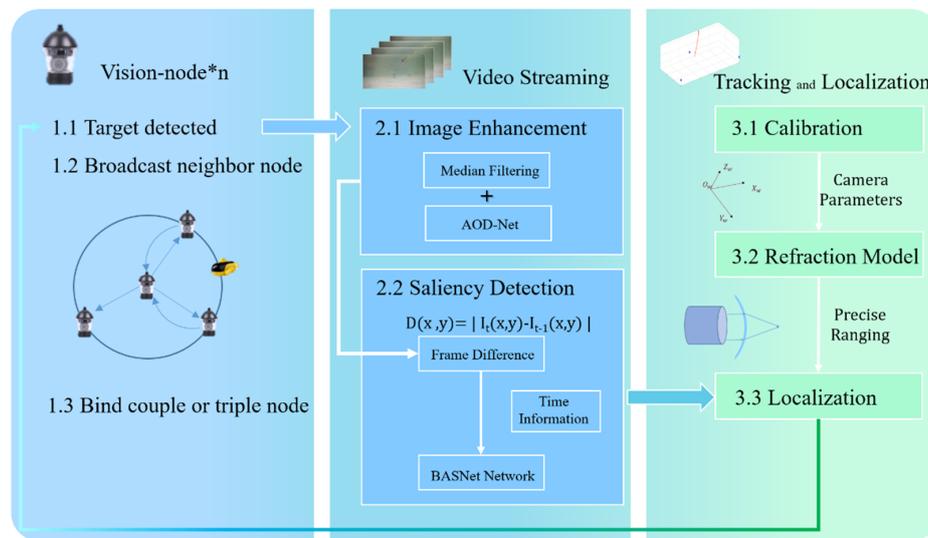


Figure 1. System flow chart.

### 3. Saliency Detection Based on Video

#### 3.1. Image Preprocessing

Compared with images on land which are clear and easily available, underwater images have serious degradation problems, which affect their reliability in underwater applications [16–18]. Because the medium in the underwater environment is liquid, and most of them are in a turbid state. The turbidity of water indicates the degree of obstruction of light transmission by suspended solids and colloidal impurities in the water. There are two main factors for underwater image degradation. The first is scattering. The light in the underwater scene will collide with suspended particles in some medium in the water before reaching the camera, and the direction of the light will change, resulting in blurred images, low contrast, and foggy effects. The second is the attenuation of light caused by the absorption of suspended particles, which depends on the wavelength of the light. Different wavelengths of light attenuate at different rates in the water, and this uneven attenuation causes color deviation of underwater images. For example, underwater images are generally dominated by blue-green tones, because longer-wavelength red light absorbs more than green and blue light. This effect has an impact on the quality of underwater images, which is mainly reflected in image color deviation, reduced contrast, low noise and low visibility [19,20]. All of these may result in reduced image quality, reduced contrast or loss of detail, and affect the detection and tracking algorithms of image targets. Therefore, the two goals of underwater image enhancement are mainly to improve image clarity and eliminate color deviation.

At present, there are many methods for image enhancement to improve the distinction between object and background. One is to use traditional image enhancement, such as histogram equalization (histogram equalization, HE), on this basis, further developed the contrast limited histogram

equalization (contrast limited adaptive histogram equalization, CLAHE), which can be directly applied to the underwater environment. There are also studies that have applied methods based on the color constancy theory (Retinex); and some that have made proposals based on the dark channel priority (DCP) algorithm and its improved algorithm. Some scholars also applied the probability-based method (probability-based method, PB), based on the probability method, to estimate the illumination and reflection in the linear domain as a way to enhance the image.

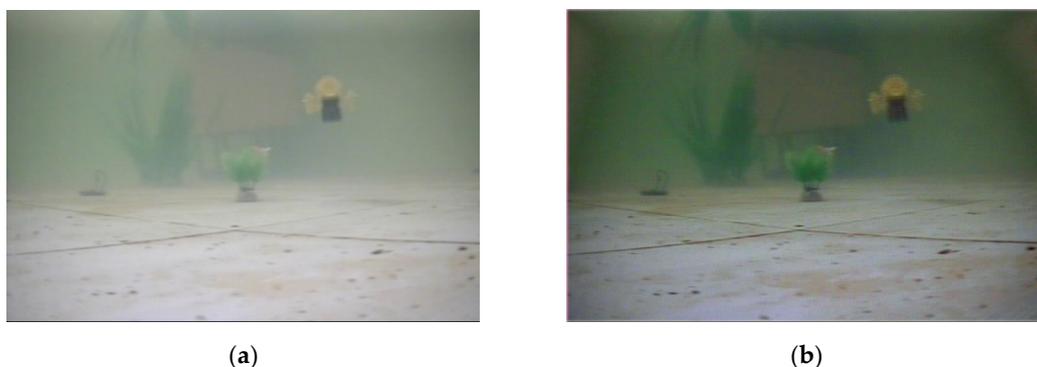
This study uses the PSNR(Peak Signal to Noise Ratio) indicator evaluation method first [21], namely peak signal to noise ratio, to compare the three basic filters. After conducting a comparative experiment, it was found that in a turbid underwater environment, the PSNR value after median filter processing is the highest, so the median filter is used to preprocess the image. The PSNR table of the three methods is displayed below. From Table 1, it can be seen that the median filter performs better than the other two filtering methods:

**Table 1.** PSNR of the three methods (PSNR results).

Method	Median Filter	Mean Filter	Gaussian Filter
PSNR	49.1133	32.2807	34.7462

The underwater image enhancement is not to improve the quality of each pixel in the image, but to improve the quality of the most important, useful, and task-related information and suppress the information of areas of no interest. Traditional image enhancement algorithms directly applied to the underwater environment often have limited enhancement effects and issues of image distortion. The reason is that these algorithms can only deal with the decrease of image contrast, but the colorcast problem of the image cannot be effectively solved. Therefore, it can be said that traditional image enhancement algorithms cannot significantly help target detection.

In addition, these algorithms cannot change according to different application scenarios due to their simple models, which often leads to excessive or poor enhancement. Therefore, underwater image enhancement, underwater target detection and tracking methods influence each other and assist in gains. Since the experiment selected target detection under started to be conducted under turbid conditions, the AOD-Net image-dehazing algorithm was innovatively applied to underwater scenarios [22]. This model is designed based on a re-formulated atmospheric scattering model, consistent with the turbid underwater scattering environment. Unlike most previous models that estimate the transmission matrix and atmospheric light separately, AOD-Net directly generates clean images through a lightweight CNN to meet the needs of turbid environments and poor computing capabilities of terminal equipment. The effects can be seen in the Figure 2.



**Figure 2.** (a) The original optical image; (b) the image after enhancement.

From the renderings, it is found that AOD-Net is very friendly to the turbid underwater environment, and the quality of underwater images is significantly improved. Such feature provided high-quality images for later target detection and tracking.

### 3.2. Saliency Detection

The image target detection algorithm is an important part of computer vision, and it has gradually developed from the traditional manual feature detection method to the detection algorithm based on deep learning. Most recent successful target detection methods are based on convolutional neural networks (CNN), and other researchers have already designed many different layouts on this basis. Wei Li et al., for instance, designed three major units: adaptive channel attention unit, adaptive spatial attention unit and adaptive domain attention unit, combining YOLOv3 and MobileNetv2 frameworks to achieve good results. In the same direction, Guennouni et al. [23] implemented a simultaneous object detection system based on local edge orientation histogram (EOH), as a feature extraction method with a small target image database. It is worth pointing out that this detection method requires a large number of target image databases. Ue-Hwan Kim et al. [24] proposed a neural architecture that can simultaneously perform geometric and semantic tasks in a single thread: synchronous visual odometer, object detection and instance segmentation (SimVODIS). Pranav Venuprasad [25] and others researchers in the field applied a deep learning model to detect the video captured by the world-view camera of the glasses to detect objects in the view field. It was found that the clustering method improves the accuracy of the appearance detection of the test video, and it is more robust on videos with poor calibration effects [26].

Due to the uneven distribution of underwater light, there are not a lot of researches available about underwater saliency. It is worth mentioning that the practice is also challenging. To design the model, this article refers to the BASNet framework proposed by Xuebin Qin and other scholars, a combination of encoder and decoder networks, residual networks, Unet and other network structure models. To detect the saliency of a single image, a network model is used to make a preliminary prediction and then another network model is used to refine the contour edges [27]. However, because this method does not distinguish between dynamic and static multi-targets in the image, this study also takes some time to improve the BASNet model.

Traditional target detection methods require feature extraction of the target. However, in the underwater monitoring network, the shape of the target entering the area to be detected cannot be known in advance, so we use the method of saliency detection for underwater monitoring. Saliency detection refers to automatically processing the region of interest when facing a scene and selectively ignoring the region of interest. BASNet is not well used for saliency detection based on video streams, because it does not consider the influence of excluding static objects with high saliency on the detection results. Therefore, we have improved the BASNet network model in this article. In the BASNet model, we chose the frame difference method with higher real-time performance to obtain the time information of the target movement, and introduced it into BASNet.

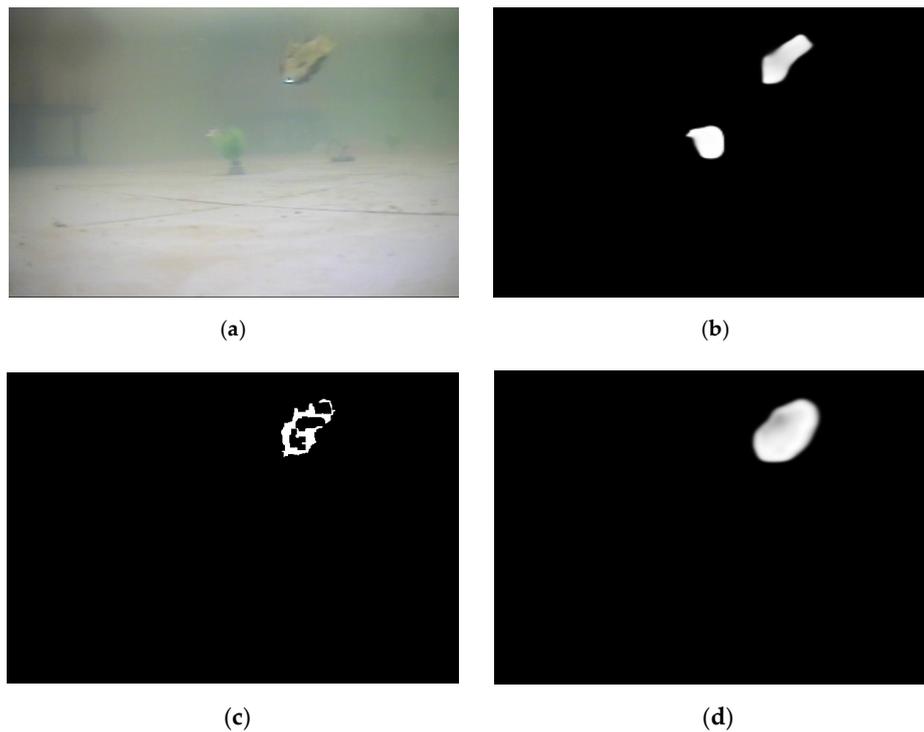
The frame difference method uses the different positions of the target in different image frames to perform difference calculation on two or three consecutive frames in time to obtain the contour of the moving target.

$$D(x, y) = |I_t(x, y) - I_{t-1}(x, y)| \quad (1)$$

$D(x, y)$  represents the difference image between two consecutive images,  $I_t$  and  $I_{t-1}$  are the grayscale images at  $t$  and  $t-1$  respectively.

$$F(x, y) = \begin{cases} 255, & \text{if } D(x, y) > T \\ 0, & \text{others} \end{cases} \quad (2)$$

where  $T$  stands for the threshold of the difference image, and  $F(x, y)$  is the image after the inter-frame difference binarization. The effects of BASNet, frame difference method and OURS can be seen in the Figure 3.



**Figure 3.** Results of single frame detection. (a) The original image; (b) result of BASNet; (c) result of frame difference method; (d) result of OURS.

Take the above group of pictures as an example. Among them, Figure 3a is the original image. Where Figure 3b, c and d are the results of detection binary map generated by BASNet, frame difference method, and OURS respectively. It can be clearly seen that the static aquatic model used as the reference coordinate will interfere in the tracking results during processing. The model detected by BASNet contains information about unwanted static objects. And the frame difference method have the potential to cause internal voids. This study first introduces time information into the BASNet framework to assist in target detection. The time information is proposed by Wenguan Wang and applied to the deep learning model used to effectively detect areas in the video [28]. Such a process is conducted as a way to innovatively use the outcomes of the frame difference method as input and, at the same time, eliminate static interference items during multi-target detection in the video stream, redirecting the attention to the dynamic AUV.

#### 4. Tracking

In order to extract the pixel position of the target in the image more clearly, enhancement processing and saliency detection methods have been conducted to solve the problem of image degradation under turbid underwater environmental conditions. Previous studies that examined the relationship between image and localization, correction, pose estimation and geometric reconstruction found out that both are important steps that need to be considered and performed [28–32].

When it comes to computer vision, the pinhole imaging model is the most common camera model. In order to get the world coordinates of the target, the first step is to obtain the camera pixel coordinates, and convert the camera pixel coordinates into the camera physical coordinates through the internal parameter matrix. Only after that it is possible to convert the camera's physical coordinates to world coordinates by the rotation matrix. The calculation method is based on the basic flow of the structure from motion (SFM) algorithm. In order to obtain the internal parameter matrix of the camera, a calibration board has to be used together with the algorithm of feature point extraction and feature point matching to extract the corner points in the multiple calibration board images as a way to calculate the internal parameter matrix.

The world coordinates of the target pixel can be converted by the following formula:

$$Z \times \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \times (R \times \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T) \quad (3)$$

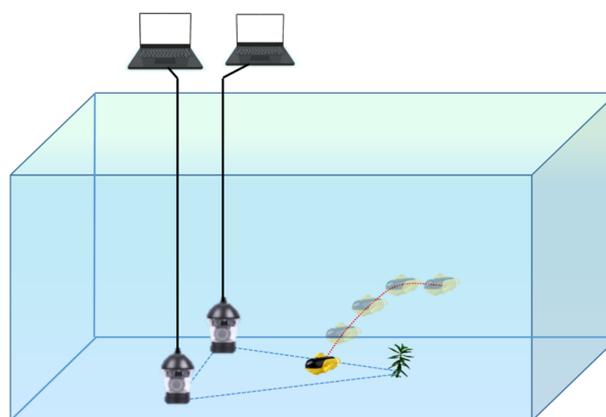
where  $Z$  is the distance from the target to the camera,  $K$  is the internal parameter matrix,  $R$  is the rotation matrix, and  $T$  represents the translation matrix.  $u$  and  $v$  stand for the camera pixel coordinates of the target, and  $X_w, Y_w, Z_w$  are the world coordinates of the target.

It is noteworthy that underwater camera requires a special waterproof cover for waterproofing. The light in the light field to be projected on the CCD (Charge Coupled Device) sensor in the camera do not only passes through the turbid water (liquid) medium, but also goes through the transparent solid medium. It also undergoes the air (gas) medium between the protective cover and the camera lens, and finally passes through the lens again to reach the sensor. Therefore, ordinary perspective models cannot be applied in the conversion process between underwater cameras and the world coordinate system. In this case, it is necessary to estimate the parameters between the camera and its housing in the refraction calibration part, and then use the modeled camera system to eliminate the refraction effect in these steps.

Further on, classic SFM algorithm can calculate the world coordinates of the target pixel after the camera calibration in the water, but the coordinate value has a certain deviation compared with the outcomes after the calibration in the air. Two sets of calibration data can be obtained without changing the positions of the camera and the calibration plate in the two media. The corresponding relationship between the two sets of corner points is obtained by calculation, and the rotation matrix  $R'$  and the translation matrix  $T'$  are updated. The next step is to use the new matrix to calculate the three-dimensional coordinate points of the target. Finally, the conversion matrix between the image coordinate system and the real coordinate system could be obtained, allowing the target to be tracked and located.

## 5. Experiment Description and Results

The experiment was carried out in a pool of 3 m × 4 m × 1 m. Camera L and Camera R were equipped with rotatable heads, placed at the bottom of the pool. For the convenience of experimental detection, the cameras were respectively pointed at the test water area at an inclination angle of 45°, and the water plants were located at the center of the central axis, revealed on the test core water area. One of the key starting processes was to drive the small underwater robot to perform random movement, and set other angle cameras for position recording. The design of system can be seen in Figure 4.



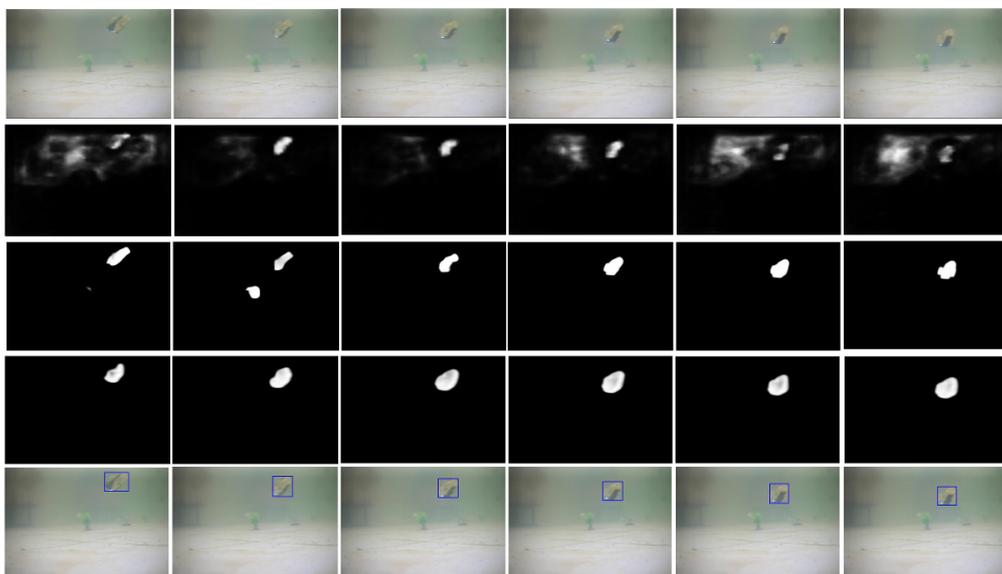
**Figure 4.** UVNets: an underwater vision target detection and tracking system.

The parameter list of experiment is in Table 2.

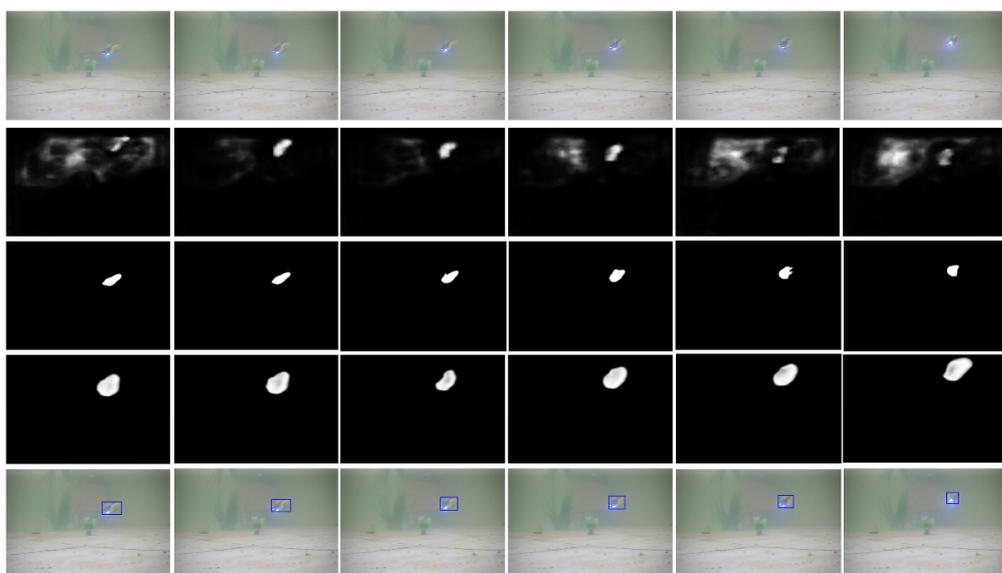
**Table 2.** Parameter list.

Name	Configuration
Pool	3 m × 4 m × 1 m
Medium	Water without salinity, more turbid
Camera L	Rotatable, with protective cover
Camera R	Rotatable, with protective cover
Calibration board	GP290
Environment	Python 3.6, Pytorch 0.4.0

The results of the saliency detection experiment can be closely seen in Figures 5 and 6, as displayed below:



**Figure 5.** Results of Camera L saliency detection.



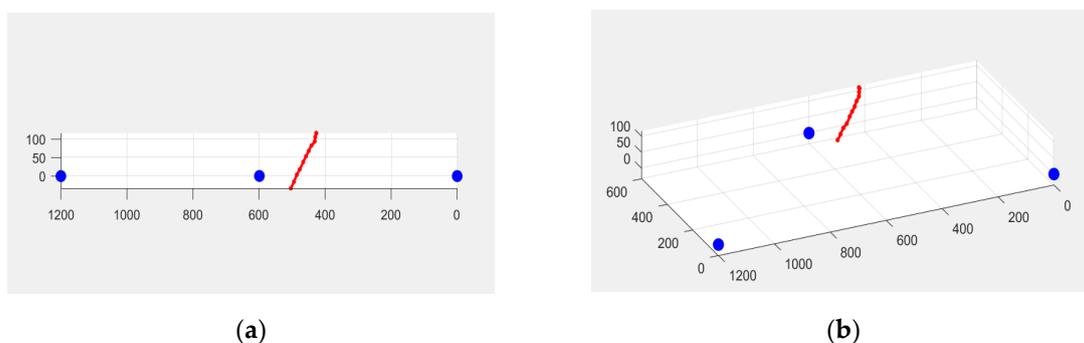
**Figure 6.** Results of Camera R saliency detection.

The group of pictures above represents eight shots taken from the camera video stream, being the first line is the original one. The second row is the dynamic detection results of the video saliency detection studied in the paper [13]. The network model based on the vgg network is used to detect the saliency of a static single frame image. In this case, the detection result and the pictures of two adjacent frames are used as input. The network structure of the dynamic video saliency detection is the same as the static one, and spatiotemporal information is introduced as part of the process. Due to the underwater lighting issue, a foggy effect happens in the binary image of the second row of the underwater image experiment, that is, the light on the water surface. However, many "fogging" phenomena occur in the entire video stream, and this part cannot be filtered in the saliency. Therefore, this saliency detection method could be used to detect underwater targets, especially close to the water surface where the light is scattered and unclear.

The third behavior is related to the binarization results after applying BASNet for saliency detection. It can be clearly seen that, in addition to identifying the normal moving AUV dynamic target in the second picture in Figure 5, the fixed plants model in the underwater area is confused as the target.

The fourth row represents the binary images and real-time tracking block diagrams of the saliency detection used in this article. It is easy to notice that the edges of the target are clearer, and only the constantly moving targets that are needed in the image are extracted. It doesn't pay attention to the objects that have no use (e.g., fixed plants) as the reference coordinates in the underwater area.

In order to ensure the accuracy of underwater ranging and localization, for the purpose of the pool experiment, the L and R cameras were calibrated under the water using the calibration board, and then the refraction model was used to analyze the target and the precise distance, and the L and R cameras were used to compare with each other. The target ray in the image determines the true position of the target. To facilitate the experiment, the experiment uses an AUV model that can float up, dive down, and simply turn. The trajectory tracking experiment results of the AUV model are as shown in Figure 7.



**Figure 7.** (a) Front view of the track; (b) 3D trajectory diagram.

The two blue dots on the left and right in Figure 7b are the position coordinates of Camera L and Camera R, while the blue dot in the middle stands for the fixed plant model as a reference for the central axis. The red line represents the moving track of the target. For a matter of interpretation, the picture on the right displays the target's trajectory in three-dimensional coordinates.

Since it is difficult to obtain the correct position of the target in real time, in order to better reflect the performance of the system, we built a linear moving device to calculate the average error of the system. The device is mainly composed of a plain shaft, a bearing, and a carrier, and the carrier can perform linear motion with low friction on the plain shaft. The structure can be seen in Figure 8.

We marked the scale on the plain shaft, so that we can get the correct position and the calculated position of the target in the picture at the same time. We conducted five sets of experiments with this set of equipment, and obtained 25 sets of coordinate points, and then obtained the average error of the

system. We have selected two sets of points and presented them in the table. The unit of measurement is millimeter.



Figure 8. Linear moving device.

The coordinate points list of experiment is in Table 3.

Table 3. Coordinate points list (Two sets of coordinate points).

Name	Point 1	Point 2	Point 3	Point 4	Point 5
Group 1	(587.86,492.7,54.40)	(586.25,542.47,50.17)	(581.91,593.18,52.21)	(577.36,642.65,63.20)	(576.09,690.11,50.51)
Group 1-origin	(586.5,500,61.5)	(586.5,550,61.5)	(586.5,600,61.5)	(586.5,650,61.5)	(586.5,700,61.5)
Group 2	(587.15,533.26,57.33)	(585.07,561.16,54.65)	(581.91,593.18,52.20)	(579.62,62.21,52.50)	(577.21,653.56,51.20)
Group 2-origin	(586.5,540,61.5)	(586.5,570,61.5)	(586.5,600,61.5)	(586.5,630,61.5)	(586.5,660,61.5)

After calculating the coordinate points, we have obtained the average error in three dimensions in Table 4. The unit of measurement is millimeter.

Table 4. Average error of the system (The average error in three dimensions).

Name	X	Y	Z
Average error	12.1702	10.2703	11.0518

The actual curve of the target movement and the calculated curve are shown in Figure 9. It can be seen that the experimental effect is in line with our expectations, and the average system error is small. According to our analysis, the error may come from the camera position error, lens correction error and manual measurement error.

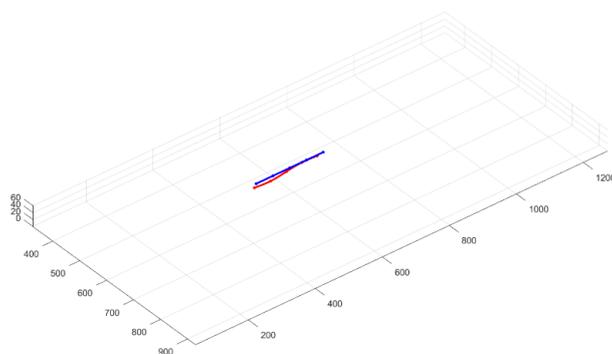


Figure 9. The actual curve (blue) and the calculated curve (red) of the target movement.

## 6. Future

As the deployment demand of visual detection nodes in underwater detection networks increases, the formation of underwater transmission nodes and observation nodes will also become denser as a consequence. At the same time, since most detection tasks require coordinated detection by mobile AUVs, the underwater sensor network will inevitably face the problem of random access of AUVs at the MAC (Media Access Control) and routing layers, especially regarding detection and distance perception issues of random access nodes. Since the location information of the node is unknown and constantly changing, maintaining its connection with the entire network will reserve a certain communication bandwidth or consume more control information to establish such communication. If the distance perception and positioning problem of this experiment is combined with the communication and access problem of the sensor network, it will be easier and more convenient for random nodes to move underwater and dynamically access or leave the network. In the future, the experiment will expand the scale of nodes and further expand the research based on the communication characteristics of underwater acoustic network. Since the power consumption of visible nodes cannot be ignored, we will also consider the energy consumption of network nodes in future research. In addition, we will also consider the impact of underwater transparency, turbidity and system error on the detection distance to enhance the stability of the underwater acoustic network for target detection.

**Author Contributions:** Conceptualization, J.L. (Jun Liu), S.G. and W.G.; methodology, J.L. (Jun Liu) and S.G.; supervision, J.L. (Jun Liu) and W.G.; writing—original draft, S.G., W.G. and H.L.; writing—review and editing, S.G., W.G., B.L., H.L. and J.L. (Jiaxin Liu). All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grants No. 61971206, No.61631008, and No. U1813217), the Fundamental Research Funds for the Central Universities 2017TD-18, and the National Key Basic Research Program (2018YFC1405800).

**Conflicts of Interest:** The authors declare no conflict of interest while producing this piece.

## References

1. Liu, J.; Du, X.; Cui, J.; Pan, M. Task-oriented Intelligent Networking Architecture for Space-Air-Ground-Aqua Integrated Network. *IEEE Internet Things J.* **2020**, *7*, 5345–5358. [[CrossRef](#)]
2. Liu, J.; Yu, M.; Wang, X.; Liu, Y.; Wei, X.; Cui, J. An underwater reliable energy-efficient cross-layer routing protocol. *Sensors* **2018**, *12*, 4148. [[CrossRef](#)] [[PubMed](#)]
3. Wei, X.; Liu, Y.; Gao, S.; Wang, X.; Yue, H. An RNN-based delay-guaranteed monitoring framework in underwater wireless sensor networks. *IEEE Access* **2019**, *7*, 25959–25971. [[CrossRef](#)]
4. Gou, Y.; Zhang, T.; Liu, J.; Wei, L.; Cui, J.H. DeepOcean: A General Deep Learning Framework for Spatio-Temporal Ocean Sensing Data Prediction. *IEEE Access* **2020**, *8*, 79192–79202. [[CrossRef](#)]
5. Wang, X.; Wei, D.; Wei, X.; Cui, J.; Pan, M. Has4: A heuristic adaptive sink sensor set selection for underwater auv-aid data gathering algorithm. *Sensors* **2018**, *12*, 4110. [[CrossRef](#)] [[PubMed](#)]
6. Liu, Y.; Wei, X.; Li, L.; Wang, X. Energy-Efficient Approximate Data Collection and BP-Based Reconstruction in UWSNs. In Proceedings of the International Conference on Smart Computing and Communication, Birmingham, UK, 11–13 October 2019.
7. Erol, M.; Oktug, S. Localization in underwater sensor networks. In Proceedings of the IEEE Signal Processing and Communications Applications (SIU), Antalya, Turkey, 17–19 April 2006.
8. Al-Salti, F.; Alzeidi, N.; Day, K. Localization Schemes for Underwater Wireless Sensor Networks: Survey. *Int. J. Comput. Netw. Commun.* **2020**, *12*, 113–130. [[CrossRef](#)]
9. Song, S.; Liu, J.; Guo, J.; Wang, J.; Xie, Y.; Cui, J.H. Neural-network based AUV Navigation for Fast-changing Environments. *IEEE Internet Things J.* **2020**, *2020*, 1. [[CrossRef](#)]
10. Wei, X.; Wang, X.; Bai, X.; Bai, S.; Liu, J. Autonomous underwater vehicles localization in mobile underwater networks. *Int. J. Sens. Netw.* **2017**, *23*, 61–71. [[CrossRef](#)]
11. Kottege, N.; Zimmer, U.R. Underwater Acoustic Localization. *J. Field Robot.* **2011**, *28*, 40–69. [[CrossRef](#)]
12. Lee, K.C.; Ou, J.S.; Huang, M.C. Underwater acoustic localization by principal components analyses based probabilistic approach. *Appl. Acoust.* **2009**, *70*, 1168–1174. [[CrossRef](#)]

13. Wang, W.; Shen, J.; Shao, L. Video Salient Object Detection via Fully Convolutional Networks. *IEEE Trans. Image Process.* **2017**, *27*, 38–49. [[CrossRef](#)]
14. Kim, J. Cooperative Localization and Unknown Currents Estimation Using Multiple Autonomous Underwater Vehicles. *IEEE Robot. Autom. Lett.* **2020**, *5*, 2365–2371. [[CrossRef](#)]
15. Wang, C.; Wang, B.; Deng, Z.; Fu, M. A Delaunay Triangulation Based Matching Area Selection Algorithm for Underwater Gravity-Aided Inertial Navigation. *IEEE/ASME Trans. Mechatron.* **2020**, *1*.
16. Liu, J.; Li, B.; Guan, W.; Gong, S. A Scale-Adaptive Matching Algorithm for Underwater Acoustic and Optical Images. *Sensors* **2020**, *20*, 4226. [[CrossRef](#)] [[PubMed](#)]
17. Wang, Y.; Ma, X.; Wang, J.; Wang, H. Pseudo-3D Vision-Inertia Based Underwater Self-Localization for AUVs. *IEEE Trans. Veh. Technol.* **2020**, *69*, 7895–7907. [[CrossRef](#)]
18. Yasukawa, S.; Nishida, Y.; Ahn, J.; Sonoda, T. Field Experiments of Underwater Image Transmission for AUV. In Proceedings of the International Conference on Artificial Life and Robotics, Oita, Japan, 9–12 January 2020.
19. Banerjee, J.; Ray, R.; Vadali, S.R.K.; Shome, S.N.; Nandy, S. Real-time underwater image enhancement: An improved approach for imaging with AUV-150. *Sadhana* **2016**, *41*, 225–238. [[CrossRef](#)]
20. Soni, O.K.; Kumare, J.S. A Survey on Underwater Images Enhancement Techniques. In Proceedings of the IEEE 9th International Conference on Communication Systems and Network Technologies, Gwalior, India, 10–12 April 2020.
21. Winkler, S.; Mohandas, P. The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics. *IEEE Trans. Broadcasting* **2008**, *54*, 660–668. [[CrossRef](#)]
22. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. AOD-Net: All-In-One Dehazing Network. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4780–4788.
23. Li, W.; Liu, K.; Zhang, L.; Cheng, F. Object detection based on an adaptive attention mechanism. *Sci. Rep.* **2020**, *10*, 1–13. [[CrossRef](#)]
24. Guennouni, S.; Mansouri, A.; Ali, A. Performance Evaluation of Edge Orientation Histograms based system for real-time object detection in two separate platforms. *Recent Pat. Comput. Sci.* **2019**, *13*, 86–90. [[CrossRef](#)]
25. Kim, U.H.; Kim, S.; Kim, J.H. SimVODIS: Simultaneous Visual Odometry, Object Detection, and Instance Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *99*, 1. [[CrossRef](#)]
26. Venuprasad, P.; Xu, L.; Huang, E.; Gilman, A.; Leanne, C.; Cosman, P. Analyzing Gaze Behavior Using Object Detection and Unsupervised Clustering. In Proceedings of the ETRA '20: 2020 Symposium on Eye Tracking Research and Applications, Stuttgart, Germany, 2–5 June 2020.
27. Qin, X.; Zhang, Z.; Huang, C.; Gao, C.; Dehghan, M.; Jagersand, M. BASNet: Boundary-Aware Salient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
28. Qiao, X.; Ji, Y.; Yamashita, A.; Asama, H. Structure from Motion of Underwater Scenes Considering Image Degradation and Refraction. *IFAC Pap. Online* **2019**, *52*, 78–82. [[CrossRef](#)]
29. Qiao, X.; Yamashita, A.; Asama, H. Underwater Structure from Motion for Cameras Under Refractive Surfaces. *J. Robot. Mechatron.* **2019**, *31*, 603–611. [[CrossRef](#)]
30. Zhuang, S.; Zhang, X.; Tu, D.; Xie, L.A. A standard expression of underwater binocular vision for stereo matching. *Meas. Sci. Technol.* **2020**, *31*, 115012. [[CrossRef](#)]
31. Chadebecq, F.; Vasconcelos, F.; Lacher, R.; Maneas, E.; Desjardins, A.; Ourselin, S.; Vercauteren, T.; Stoyanov, D. Refractive Two-View Reconstruction for Underwater 3D Vision. *Int. J. Comput. Vis.* **2019**, *128*, 1101–1117. [[CrossRef](#)]
32. Kong, S.; Fang, X.; Chen, X.; Wu, Z.; Yu, J. A NSGA-II-Based Calibration Algorithm for Underwater Binocular Vision Measurement System. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 794–803. [[CrossRef](#)]

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



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).