



## Article Deep Learning-Based Docking Scheme for Autonomous Underwater Vehicles with an Omnidirectional Rotating Optical Beacon

Yiyang Li <sup>1,2</sup>, Kai Sun <sup>1,\*</sup>, Zekai Han <sup>1,2</sup> and Jichao Lang <sup>1,2</sup>

- State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; liyiyang@sia.cn (Y.L.); hanzekai@sia.cn (Z.H.); langjichao@sia.cn (J.L.)
- <sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** Visual recognition and localization of underwater optical beacons are critical for AUV docking, but traditional beacons are limited by fixed directionality and light attenuation in water. To extend the range of optical docking, this study designs a novel omnidirectional rotating optical beacon that provides 360-degree light coverage over 45 m, improving beacon detection probability through synchronized scanning. Addressing the challenges of light centroid detection, we introduce a parallel deep learning detection algorithm based on an improved YOLOv8-pose model. Initially, an underwater optical beacon dataset encompassing various light patterns was constructed. Subsequently, the network was optimized by incorporating a small detection head, implementing dynamic convolution and receptive-field attention convolution for single-stage multi-scale localization. A post-processing method based on keypoint joint IoU matching was proposed to filter redundant detections. The algorithm achieved 93.9% AP at 36.5 FPS, with at least a 5.8% increase in detection accuracy over existing methods. Moreover, a light-source-based measurement method was developed to accurately detect the beacon's orientation. Experimental results indicate that this scheme can achieve high-precision omnidirectional guidance with azimuth and pose estimation errors of -4.54° and 3.09°, respectively, providing a reliable solution for long-range and large-scale optical docking.

**Keywords:** underwater optical beacon; docking technology; pose detection; deep learning; underwater localization

## 1. Introduction

Autonomous underwater vehicles (AUVs), a subset of unmanned underwater vehicles (UUVs), have long been a focal point of research within underwater robotics. They are extensively used in marine science research, ocean resource surveys, and maritime security. In marine science, AUVs facilitate deep-sea mapping, current tracking, and ecosystem studies, generating high-resolution data critical for understanding climate change and biodiversity [1]. In resource exploration, AUVs enable the detailed mapping of underwater geological features and the assessment of mineral and hydrocarbon deposits, reaching depths and terrains beyond human capability [2]. Additionally, they support critical tasks in maritime security and surveillance, providing valuable data in areas that are challenging or hazardous for human divers to access [3]. Effective guidance and docking systems of AUVs are vital, as they enable routine tasks such as charging, data offloading, and maintenance to be performed autonomously, reducing the need for human intervention and extending mission duration. In modern underwater guidance and docking systems, optical beacons serve as highly efficient guidance tools, playing a crucial role, especially in achieving high-precision autonomous guidance and docking.

Underwater optical docking technology employs underwater cameras to identify optical beacons such as LEDs and laser diodes (LDs) to guide the docking of AUVs with base



Citation: Li, Y.; Sun, K.; Han, Z.; Lang, J. Deep Learning-Based Docking Scheme for Autonomous Underwater Vehicles with an Omnidirectional Rotating Optical Beacon. *Drones* **2024**, *8*, 697. https://doi.org/10.3390/ drones8120697

Academic Editor: Giancarlo Rufino

Received: 15 October 2024 Revised: 5 November 2024 Accepted: 19 November 2024 Published: 21 November 2024



**Copyright:** © 2024 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/).

<sup>\*</sup> Correspondence: sunk@sia.cn

stations. Particularly in close-range docking processes, optical positioning technologies provide extremely high positional accuracy and stability for AUVs, thus becoming the primary means for precise short-range underwater docking [4]. Key indicators of underwater optical docking include range of operation, positioning accuracy, and update frequency. In recent years, advancements in computer vision technology and hardware upgrades have enhanced the positioning accuracy of optical methods to centimeter-level, with real-time updates being possible. Specifically, Zhang et al. [5,6] achieved AUV docking using an L-shaped light array, controlling the positioning errors within 20 cm horizontally, 15 cm longitudinally, and 15 cm vertically. Petar et al. [7] designed an optical docking system using four asymmetrically arranged flashing LEDs, achieving a positioning accuracy of 20 cm within 4 m. Cheng et al. [8] proposes a real-time method based on polarized optical guidance using four polarized artificial underwater landmarks with a localization position error of no more than 0.116 m. Zhao et al. [9] introduced a dual-type marker fusion-based underwater visual positioning method, combining light sources and ArUco markers, achieving positioning accuracies between 1.62 cm and 2.39 cm. Furthermore, Sun et al. [10] utilized a 460 nm blue LED beacon for close-range guidance and docking within a 15 m range. Zhejiang University [11] used a 94w power white SLS-5200 underwater lamp combined with a line of sight (LOS) guidance scheme to complete docking over a distance of 20–30 m on a lake. In 2024, they [12] utilized an improved detection algorithm to achieve wide-range monocular single-light visual guidance. Cai et al. [13] addresses the docking of the autonomous underwater helicopter (AUH) by installing two blue guide lights with large diffusion angles on the landing platform and realizing high-precision localization through binocular two-light vision algorithms. Dörner et al. [14] realized a six-degree-offreedom attitude reconstruction of an AUV at a range of 7 m using a set of optical beacons with anterior and posterior height differences. Xu et al. [15] proposed a stereoscopic vision navigation method using four green LEDs, achieving centimeter-level positioning accuracy within a maximum range of 20 m. The Institute of Semiconductors [16,17] developed a blue laser diode docking light named "Beijixing," which can recognize up to 18 m. Furthermore, they employed intersecting laser lines to extract four-point correspondences, successfully achieving optical positioning over a distance of 10 m in a controlled pool environment. Chen et al. [18] designed a node with integrated laser guidance and communication capabilities, and this laser docking system is capable of guiding AUVs to communicate with the seafloor observation network.

Despite advancements in optical docking precision, existing solutions still face considerable challenges regarding their operational range and distance. The propagation of light through water is significantly affected by absorption and scattering, which typically restricts the emission angles of LED and LD sources. While increasing the emission angle can extend the capture range, overly broad angles lead to a substantial loss in light energy, thereby reducing the guiding distance. Consequently, current optical beacons must strike a balance between emission angles and transmission distances, typically limiting their operational range to approximately 20 m within a sector-shaped area. The constrained range in both distance and direction significantly diminishes the success rate of AUV docking, as it heavily relies on the accuracy of the return navigation phase, resulting in low fault tolerance. Against this background, this paper introduces an innovative design for an omnidirectional rotating optical beacon. Utilizing 360-degree dynamic scanning, the beacon significantly enhances docking range and capture scope, effectively addressing directional capture limitations in existing systems. While extending the distance, it provides precise multi-angle guidance for the AUV, facilitating smooth adjustments and successful docking despite directional deviations during the return phase, thereby markedly improving mission success rates. This omnidirectional rotating optical beacon aligns with current trends in intelligence and autonomy, offering essential technical support for extendedrange guidance and multi-angle docking in complex underwater environments for future AUV applications.

The detection of underwater optical beacons is another critical issue for the optical docking of AUVs. Water scattering and environmental noise cause significant variations in the size and shape of optical beacons during docking, making accurate localization of the beacon's center challenging and impacting the precision of optical docking. Currently, the methods for detecting underwater optical beacons primarily fall into three categories: traditional image feature extraction, a hybrid approach combining deep learning-based object detection with traditional optical center feature extraction, and optical center detection based solely on deep learning. Traditional image processing methods often utilize color and brightness characteristics to highlight the optical beacon, combining adaptive threshold segmentation and morphological fitting to extract features of the center [15,19–21]. Although straightforward, these methods lack adaptability to changes between the light source and background, exhibit limited robustness, and achieve lower detection precision. Deep learning-based object detection techniques, such as YOLO and Fast R-CNN, can extract deeper features of the optical beacon, directly detecting the entire docking station or individual beacons and then utilizing traditional image processing to extract the optical center in the detected areas [22,23]. Despite their excellent robustness and detection accuracy, these two-stage processes increase the complexity of the algorithms. Moreover, traditional methods of optical center feature extraction are susceptible to changes in beacon shape, which can reduce the accuracy of center detection. In contrast, deep learning algorithms treat localization as a regression problem, directly outputting the optical center position from input images using convolutional neural networks [24]. This approach is structurally simpler and more efficient in detection but has weaker interference resistance and is prone to misidentification.

To address the issues outlined above, this study proposes a deep learning-based approach for simultaneous light source detection and centroid extraction. We implemented and refined this approach using the YOLOv8-pose model specifically for underwater optical beacon scenarios. The method processes object detection and centroid extraction tasks in parallel, improving detection speed compared to traditional image processing and deep learning methods that rely on conventional feature extraction for centroid determination. Furthermore, it ensures accurate centroid localization even when the light source undergoes deformation, thereby enhancing the precision of optical center positioning. Compared to direct deep learning centroid detection methods, our algorithm leverages detection bounding boxes during inference to help filter out the correct centroids, significantly reducing the false detection rate. In response to challenges posed by multi-scale light sources in underwater optical docking contexts, we have added a detection head for small targets and optimized the network through dynamic convolution and RFApose detection heads (receptive-field attention convolution integrated into pose detection heads). Lastly, we introduced a post-processing method based on keypoint and intersection over union (IoU) matching that leverages keypoint positional data to help filter detection boxes, effectively eliminating redundant matches. This research applies the proposed underwater optical beacon detection algorithm to the detection of omnidirectional rotating optical beacons and introduces a light source feature-based metric method on this basis, achieving precise detection of beacon orientation. Through synchronized scanning and LOS methods, this study utilizes the omnidirectional rotating optical beacon to obtain directional and attitudinal information about AUVs, providing support for long-distance and omnidirectional data.

The main contributions of this study are as follows:

- An underwater omnidirectional rotating optical beacon was designed to offer a 360-degree operational range of up to 45 m. The design overcomes the limitations of traditional underwater optical beacons, which are hindered by restricted directions and shorter detection distances, thus enhancing docking success rates.
- We have created an underwater optical beacon dataset with manually annotated target boxes and centroid keypoints. A deep learning-based algorithm was developed for the parallel detection of optical beacons and centroids, which is an improved YOLOv8-pose model that significantly enhances detection performance. The algo-

rithm achieved 93.9% AP at 36.5 FPS, with at least a 5.8% increase in detection accuracy over existing methods.

• For the omnidirectional rotating optical beacon, we developed a metric method based on light source features that achieves correct beacon orientation detection. Combined with synchronized scanning and LOS methods, the azimuth and pose estimation errors of this approach are 4.72° and 3.09°, respectively, which meet the practical requirements.

## 2. Underwater Omnidirectional Rotating Optical Beacon Docking System

#### 2.1. Docking Approach for the Underwater Omnidirectional Rotating Optical Beacon

In the design of the docking system for underwater omnidirectional rotating optical beacons, the core strategy involves installing these beacons on base stations to perform a 360-degree dynamic scan. This provides precise location and orientation information to AUVs from all directions, facilitating accurate guidance for docking. To address the limited range of traditional optical beacons, the beacons designed in this study feature a narrowed emission angle with focused light sources, significantly extending the transmission distance to 40–50 m at the same power output. However, this design also introduces the challenge of a reduced operational range. To overcome this, we employed a planar 360-degree rotational scanning technique, trading time for space to achieve long-range guidance in all directions. Moreover, the continuous scanning of the omnidirectional rotating optical beacon ensures uninterrupted signal coverage, reducing signal interruptions caused by direction changes and enhancing the docking process's stability. The specific setup, as shown in Figure 1, includes a docking control cabin within the docking base station, housing a timer and the omnidirectional rotating optical beacon. The AUV is equipped with an underwater camera and a vision computing board to capture and process images of the optical beacon. The docking control cabin powers the light source and the motor, controlling the motor to drive the reflector for a 360-degree light scan. After capturing the image of the optical beacon, the underwater camera of the AUV transmits it to the vision computing board for recognition and angle calculation, guiding the movement of the AUV. Before docking, the AUV and the base station's docking control cabin synchronize their timers for time alignment, and the AUV's angle is calculated based on a predetermined rotational strategy and the current time.



Figure 1. Framework of the underwater omnidirectional rotating optical beacon docking system.

As illustrated in Figure 2, the omnidirectional rotating optical beacon effectively expands the pre-adjustment range for AUVs during long-distance docking. Even if the AUV is positioned behind the docking station, it can begin directional adjustments from a distance, significantly enhancing the success rate of docking operations. At closer ranges,

the docking station uses traditional optical beacons, utilizing their fan-shaped docking areas for precise pose estimation. Consequently, this docking scheme not only demonstrates efficiency and reliability but also shows adaptability and flexibility in practical applications, which is crucial for AUVs operating long-term in complex marine environments.



Figure 2. Schematic of the underwater omnidirectional rotating optical beacon docking system.

### 2.2. Design of Underwater Omnidirectional Rotating Optical Beacon

The main structure of the underwater omnidirectional rotating optical beacon is shown in Figure 3. The beacon consists of a mounting case, a transparent protective cover, a light barrel, a reflector, and a rotating mechanism. Both the light barrel and the reflector's rotating mechanism are housed within a sealed case, with the transparent protective cover situated above the case. The reflector is affixed to an inclined bracket and connected to the rotating mechanism via a screw. The rotating mechanism includes a motor, motor gears, and drive gears, with the motor mounted inside the case to rotate the reflector via gear engagement. Positioned beneath the drive gears, the light barrel contains lenses and a light filament, with the lenses clamped between two lens mounts. Additionally, a heat sink connected to the bottom case is installed beneath the light filament to efficiently conduct heat from the electronic components. During the beacon's operation, light emitted from the light source is focused through the lenses, reflected by the rotating reflector, and then emitted through the transparent protective cover. The position of the lenses can be adjusted by modifying the fixed mounts, altering the focus of the light, which in turn affects the light's operational distance and angle. Furthermore, the angle of the reflector can be adjusted by replacing the supports and shims with different angles, thus flexibly changing the direction of the light output.



Figure 3. Structural diagram of the underwater omnidirectional rotating optical beacon.

To achieve long-distance guidance with the optical beacon and obtain optimal light source characteristics, we carefully selected the light source. According to the research by Duntley S.Q. [25], blue-green light with a wavelength of 400–550 nm attenuates much less in water compared to other wavelengths. Therefore, we chose a blue LED with a wavelength of 450 nm as the light source to reduce transmission losses. To obtain better light source image information, we tested the imaging characteristics of the blue guide light under different power, distance, and beam divergence conditions in a pool environment, as illustrated in Figure 4. The results indicated that at the same power, the narrower the beam divergence angle of the light source, the greater the transmission distance. To surpass existing transmission distance limitations, we reduced the beam divergence angle to within 10 degrees at 30 W, extending the effective operational distance to 45 m while maintaining excellent directionality.



Figure 4. Underwater light source selection. (a) 10 W,  $60^{\circ}$ ; (b) 30 W,  $60^{\circ}$ ; (c) 30 W,  $10^{\circ}$ .

#### 3. Deep Learning-Based Detection Algorithm for Underwater Optical Beacon

In the complex underwater environment, spatial distribution and scattering cause light spots to distort, making it difficult to accurately extract the centroids of optical beacons and affecting positioning precision. This issue is especially pronounced when detecting omnidirectional rotating optical beacons, where light sources at different rotation angles produce varied imaging shapes, such as haloed trailing ellipses and circles. Traditional feature extraction methods struggle to effectively address these challenges. Consequently, this study introduces a deep learning algorithm for the real-time detection and centroid extraction of underwater optical beacons. Leveraging its powerful feature extraction capabilities, deep learning enhances the accuracy and stability of optical beacon detection in complex and variable underwater conditions.

#### 3.1. Underwater Optical Beacon Dataset

This study compiled a dataset of 7372 real underwater degraded images collected in pool and lake environments, encompassing various quantities and distances of optical beacons and considering factors such as blur and noise light interference. The underwater optical beacon dataset includes images of eight lights, six lights, four lights, two lights, and single light setups, with the eight-light images sourced from the UDID dataset [26] and the remaining images derived from video data from docking experiments. We recorded videos in multiple underwater environments using a high-resolution underwater camera to ensure clarity and detail, capturing footage at 30 frames per second (FPS) to maintain stability in dynamic scenes. Automated scripts were employed to extract images from the recorded videos, selecting one image every 15 frames to ensure sufficient temporal separation and avoid redundancy from consecutive frames. We randomly divided the dataset into training and validation sets at a ratio of 5:1, ensuring that images of each light quantity were distributed proportionally. Ultimately, the training set contained 6144 images, and the test set comprised 1228 images. During the data division, it was ensured that each video segment was assigned exclusively to either the training or test set to prevent overlap between consecutive frames. This division method helps consider the diversity of sample scenarios during the model training and validation phases, enhancing the model's generalization capability. Moreover, to mitigate the risk of overfitting due to the small dataset, we implemented various data augmentation techniques, including random rotation, flipping, scaling, brightness adjustment, and color transformation, thereby increasing the dataset are detailed in Table 1.

	Eight Lights	Six Lights	Four Lights	Two Lights	Single Light *
Examples	202		11		1
Training Set	1535	1374	350	1805	1080
Test Set	307	274	70	361	216
Image Size	$720 \times 576$	$960 \times 576$	$960 \times 576$	$1920 \times 1080$	$1920 \times 1080$

Table 1. Parameters of the underwater guide light dataset.

\* The single light images are from the underwater omnidirectional rotating beacon experiment.

For the annotation task of underwater optical beacons, we utilized the Labelme v4.5.9 software. First, we delineated the position of each optical beacon with bounding boxes, ensuring that these boxes tightly encompassed each light source. Then, we marked the centroid of each optical beacon as the unique keypoint within each bounding box. Statistical data reveal that our dataset comprises 26,476 annotated instances. Figure 5 illustrates the annotated visualization information of the dataset. Figure 5a displays a joint distribution histogram of the bounding box positions, revealing that while the centers of the bounding boxes predominantly cluster near the center of the images, they are distributed throughout the image. Figure 5b shows histograms of the heights and widths of the bounding boxes, demonstrating that the dimensions of these boxes are generally concentrated within half the image size. Notably, the bounding boxes are predominantly small and include not only squares but also elongated rectangles shaped by the spatial distribution of the light sources and the scattering effects underwater.

![](_page_6_Figure_7.jpeg)

**Figure 5.** Annotation information of the underwater optical beacon dataset. (**a**) Normalized positions of the bounding boxes; (**b**) Normalized sizes of the bounding boxes. Both panels are presented through histograms with 50 bins per dimension, with darker colours indicating more partitions.

In summary, the proposed underwater optical beacon dataset offers several advantages: Firstly, it covers different water quality environments, ranging from clear pool waters to algae-rich lake waters. Secondly, the dataset includes a variety of lighting conditions, from bright images in shallow water areas to dim images in deeper waters, supporting the algorithm's adaptability across different lighting scenarios. Moreover, the dataset showcases the complete morphological characteristics of underwater optical beacons from various viewing angles, not only capturing images of fixed-position optical beacons but also utilizing omnidirectional rotating optical beacons to record guide lights from multiple perspectives and attitudes. This diversity aids the algorithm in learning the high-dimensional features of optical beacons, enhancing its detection capabilities across different forms. Lastly, the dataset has undergone multiple rounds of filtering and manual annotation, providing pixel-level precise bounding box annotations, which lay a solid foundation for accurate model training and validation.

#### 3.2. Underwater Optical Beacon Detection Algorithm Based on YOLOv8-Pose

YOLOv8-pose is a deep learning algorithm designed explicitly for keypoint detection tasks, extending the classic YOLOv8 framework and enabling simultaneous object detection and keypoint recognition. Compared to methods that train object and keypoint detectors separately, this unified approach simplifies the training process and enhances training efficiency. YOLOv8-pose was selected as our preferred model for its low computational and parameter requirements while still delivering high detection accuracy and real-time performance, making it especially suitable for operation on resource-constrained devices. These characteristics make YOLOv8-pose ideal for practical engineering applications in detecting underwater optical beacons.

#### 3.2.1. Network Architecture

The network architecture of YOLOv8-pose consists of three main components: backbone, neck, and head. The backbone extracts multi-scale feature information from the input image through convolution and C2f modules. The neck employs a path aggregation network to integrate multi-level features. The head is equipped with three decoupled detection heads that target large, medium, and small objects, respectively, and calculate features for classification, object box regression, and keypoint regression tasks. However, due to the multi-scale nature of underwater optical beacon detection scenarios, the baseline model's performance could have been better. Several key improvements were made to the YOLOv8-pose model to enhance model performance in this context. First, a prediction head was added for small objects to mitigate the impact of object scale variations on detection. Second, dynamic convolution replaced the  $3 \times 3$  convolution in the C2f structure bottleneck in the neck [27] to improve the model's generalization capabilities. Lastly, receptive-field attention convolution (RFAConv) was introduced in the pose detection head [28], thereby boosting detection accuracy. The improved YOLOv8-pose architecture is illustrated in Figure 6, and the following sections will detail the enhancements made.

- Small target detection head: In optical docking, operational range is a critical metric, which is why our dataset includes numerous small object instances at long distances. Zhang et al. [29] have shown that shallower features might be more effective for such small, indistinct targets. Consequently, we introduced a specialized prediction head at the P2 layer designed specifically for detecting small targets. This quad-head structure significantly mitigates the adverse effects of substantial changes in object scale, thereby markedly improving the detection performance for small targets.
- C2f\_DC: Dynamic convolution, an extension of traditional convolutional, processes input data by dynamically selecting or combining different convolutional kernels for each input sample. It adapts to the input characteristics by adjusting parameters

through a learnable multilayer perceptron (MLP) network that generates weights controlling the contribution of each kernel. The process operates as follows:

$$\begin{cases} a = MLP(GAP(X)) \\ Y = \sum_{i=1}^{M} a_i(X * W_i) \end{cases}$$
(1)

where *X* represents the input features and *Y* the output.  $W_i$  represents the convolutional kernels, each controlled by a dynamic coefficient  $a_i$ , generated by processing the globally averaged pooled input features through a small-scale MLP network. Specifically, dynamic convolution enhances network performance under low FLOP conditions. By incorporating dynamic convolution into the C2f structure, we significantly improved detection accuracy in underwater environments. Dynamic convolution can adaptively select the optimal convolution kernels to address the varying imaging results caused by the scattering and attenuation of underwater light sources. This allows the model to better capture critical features while minimizing interference. This approach not only facilitates the deployment of more complex network architectures in resource-constrained environments but also enhances the model's generalization capability across diverse underwater scenarios.

RFApose detection head: RFAConv combines spatial attention mechanisms with convolution operations to optimize how the convolution kernels process spatial features within their receptive fields, as illustrated in Figure 7. H, W, and C in the figure represent the height, width, and number of channels of the feature map, respectively. K denotes the size of the convolution kernel. By introducing attention mechanisms, RFAConv transcends traditional spatial dimensions, enabling the network to more precisely understand and process key areas of the image. The adaptation enhances feature extraction accuracy, particularly in underwater environments characterized by low visibility and light scattering. Furthermore, it optimizes attention weights for large kernel convolutions, effectively addressing the challenge of shared kernel parameters. By reconstructing feature maps, RFAConv further enhances the encoding of image contextual information, allowing the network to better discern the relationship between noise and target light sources in underwater scenes, thereby effectively avoiding erroneous detection of interfering light. In this study, we integrated RFAConv into the decoupled detection head, enabling it to extract more precise classification, bounding box, and keypoint information from multi-scale feature maps, thus helping YOLOv8-pose more effectively address the challenge of indistinct optical beacon features caused by complex underwater environments.

![](_page_8_Figure_5.jpeg)

Figure 6. Improved network architecture of YOLOv8-pose.

![](_page_9_Figure_2.jpeg)

Figure 7. Structure of RFAConv.

#### 3.2.2. Post-Processing Based on Keypoint Joint IoU Matching

The post-processing of the YOLOv8-pose model primarily involves parsing the network's output and applying non-maximum suppression (NMS) [30]. The outputs typically include the bounding box coordinates, class confidence scores, as well as keypoint coordinates with their confidence scores. However, when processing the underwater optical beacon dataset, the scattering effects of underwater light often result in halos in the images, leading to multiple overlapping bounding boxes for the same light source target, as shown in Figure 8. These redundant detections possess similar confidence scores to actual beacons, making them difficult to filter out using simple confidence thresholds. Moreover, variations in size and proportions among overlapping targets further challenge traditional NMS methods, which rely solely on confidence and IoU thresholds and struggle to remove these redundant boxes effectively.

![](_page_9_Figure_6.jpeg)

Figure 8. Example of redundant bounding boxes.

To address this issue, this study introduces a novel NMS post-processing method based on keypoint joint IoU matching, which refines the selection of detection results by incorporating keypoint information. Specifically, we use the object keypoint similarity (OKS) [31] as a metric for keypoint information, combining it with IoU to serve as the criteria for NMS, thereby effectively removing redundant detections with high keypoint similarity. The formula for OKS is as follows, with values ranging from 0 to 1, where values closer to 1 indicate higher similarity. Here,  $d_i$  represents the distance between detected keypoints and true keypoints, s is the scale factor of the bounding box,  $k_i$  is constant

coefficients for each keypoint, and  $\delta$  ( $v_i > 0$ ) is an indicator function used to determine the visibility of keypoints.

$$OKS = \frac{\sum_{i} e^{-\frac{d_{i}^{2}}{2s^{2}k_{i}^{2}}} \delta(v_{i} > 0)}{\sum_{i} \delta(v_{i} > 0)}$$
(2)

The specific steps of this algorithm are outlined in Algorithm 1. The variables  $bbox_{det}$  and  $keypoint_{det}$  represent the bounding box and keypoint position information output by the network, respectively, while  $bbox_{filter}$  and  $keypoint_{filter}$  are the final filtered detection results. The parameters  $\lambda_{conf}$ ,  $\lambda_{iou}$ , and  $\lambda_{oks}$  correspond to the confidence threshold, the bounding box IoU threshold, and the keypoint similarity threshold, respectively. The keypoint joint IoU matching-based post-processing method effectively resolves the issue of multiple overlapping detections in underwater guide light detection, enhancing the model's accuracy in underwater target detection and providing a more reliable basis for subsequent orientation determination and pose estimation for AUVs.

Input: { $bbox_{det}$ }, { $bbox_{conf}$ }, { $keypoint_{det}$ }, $\lambda_{conf}$ , $\lambda_{iou}$ , $\lambda_{oks}$ Output: { $bbox_{filter}$ }, { $keypoint_{filter}$ }         1       Initialization         2       { $bbox_{filter}$ } \leftarrow []         3       { $bbox_{det}$ }, { $keypoint_{det}$ } ← { $bboxs$ , $keypoints   bbox_{conf} \ge \lambda_{conf}$ }
Output:       {bbox_{filter}}, {keypoint_{filter}}         1       Initialization         2       {bbox_{filter}} \leftarrow []         3       {bbox_{det}}, {keypoint_{det}} \leftarrow {bboxs, keypoints   bbox_{conf} \ge \lambda_{conf}}
1 Initialization 2 { $bbox_{filter}$ } $\leftarrow$ [] 3 { $bbox_{det}$ }, { $keypoint_{det}$ } $\leftarrow$ { $bboxs$ , $keypoints   bbox_{conf} \ge \lambda_{conf}$ }
$2  \{bbox_{filter}\} \leftarrow [] \\ 3  \{bbox_{det}\}, \{keypoint_{det}\} \leftarrow \{bboxs, keypoints \mid bbox_{conf} \ge \lambda_{conf}\}$
3 $\{bbox_{det}\}, \{keypoint_{det}\} \leftarrow \{bboxs, keypoints   bbox_{conf} \ge \lambda_{conf}\}$
4 order $\leftarrow$ sort ({bbox <sub>conf</sub> }, descending)
5 while numel (order) > 0 do
6 $i \leftarrow order$ [0]
7 $\{bbox_{filter}\} \leftarrow \{bbox_{filter}\} \cup bbox_{det} [i]$
8 {keypoint <sub>filter</sub> } $\leftarrow$ {keypoint <sub>filter</sub> } $\cup$ keypoint <sub>det</sub> [i]
9 <b>if</b> <i>numel</i> ( <i>order</i> ) = 1 <b>then break</b>
10 $\{bbox_{remian}\} \leftarrow \{bbox_{det} \ [order \ [1:]]\}$
11 $\{keypoint_{remain}\} \leftarrow \{keypoint_{det} [order [1:]]\}$
12 $\mu_{iou} \leftarrow \text{IoU} (bbox_{det} [i], \{bbox_{remian}\})$
13 $\mu_{oks} \leftarrow OKS (keypoint_{det} [i]_i, \{keypoint_{remain}\})$
14 order $\leftarrow$ order [where $\mu_{iou} < \lambda_{iou}$ and $\mu_{oks} < \lambda_{oks}$ ]
15 <b>end</b>
16 <b>return</b> { <i>bbox<sub>filter</sub></i> }, { <i>keypoint<sub>filter</sub></i> }

#### 3.3. Experiments on Underwater Optical Beacon Detection Algorithm

#### 3.3.1. Experimental Setup

Table 2 outlines the hardware and software configurations utilized in the experiments. Uniform training parameters were applied across different experimental groups during the training phase to ensure accuracy. The input image resolution was set to  $640 \times 640$  pixels, using an SGD optimizer with an initial learning rate of 0.01, momentum of 0.937, and a decay rate of 0.0005. The model was trained for up to 300 epochs with a batch size of 48, employing early stopping if no improvements were observed after 50 epochs. Inference was conducted on the Jetson AGX Orin board. The confidence threshold was set at 0.1, with a batch size of 1, indicating frame-by-frame image processing and other parameters were consistent with the training phase.

The evaluation metrics in this study included the average precision of the bounding box and keypoint detections, the number of parameters, computational workload, and FPS. For the single-category underwater optical beacon detection scenario, target detection performance was assessed using  $AP_{iou50}$  and  $AP_{iou50-95}$ , where the former indicates the average precision at an IoU threshold of 50%, and the latter aggregates average precisions across multiple IoU thresholds from 50% to 95%. The precision of detecting the optical beacon's centroid was evaluated using  $AP_{oks50}$ , representing the average precision at an OKS threshold of 50%. Additionally, FLOP measures the number of floating-point operations per image processed by the model, while the parameter indicates the total number of trainable parameters. FPS evaluates the model's real-time processing capability, a critical indicator for determining the model's practical applicability.

Environment	Specification (Train)	Specification (Inference)
CPUs	$2 \times$ Intel Xeon Gold 6234	12-core ARM Cortex-A78AE
GPU	NVIDIA RTX A6000 (48G)	NVIDIA Ampere, 2048 CUDA cores, 64 Tensor Cores
CUDA	11.3	11.4
PyTorch	1.11.0	1.13.0

Table 2. The experimental setting.

#### 3.3.2. Comparative Experiments

To validate the proposed deep learning-based optical beacon detection algorithm, we conducted comparative experiments against existing underwater optical beacon and optical center localization methods. Specifically, the performance of traditional image feature extraction methods [21], deep learning-based object detection combined with grayscale centroid methods [22], and CNN-based optical center localization [24] were assessed on the test set images.

As indicated in Table 3, our method excelled across multiple evaluation metrics and significantly outperformed traditional and other deep learning approaches. Specifically, our algorithm achieved a 10.2% improvement in detection accuracy compared to traditional image feature extraction methods, while increasing processing speed from 28 FPS to 36.5 FPS, thereby meeting real-time detection requirements. This enhancement is primarily due to the traditional methods' limited adaptability to variations in daylight and noise, challenges that our algorithm effectively addresses. Although the CNN-based method recorded the fastest detection speed at 67.9 FPS, its keypoint accuracy was only 86.7%, significantly lower than our method's 93.9%, leading to frequent misidentifications and omissions. Compared to the YOLOv8 with centroid method, our approach maintained a similar detection speed while enhancing the precision of bounding box and keypoint detection by 3.2% and 5.8%, respectively. The YOLOv8-pose baseline model, despite a slight precision decrease of 0.8% compared to the YOLOv8 centroid method, demonstrated a significant speed advantage, laying a strong foundation for future model optimization. Furthermore, our comparisons with the YOLOv9 and YOLOv10 models as target detection frameworks revealed that our model consistently maintained superior detection precision. While the YOLOv9t showed marginally better detection precision than the YOLOv8n, it fell short in detection speed, primarily due to its optimization for complex scenes, which does not confer substantial advantages for a single-target dataset like underwater optical beacons. The YOLOv10n demonstrated a 1% improvement in AP<sub>iou50</sub> compared to the YOLOv8n, but its performance declined by over 2% in AP<sub>iou50-95</sub>. This decline is attributed to the interference caused by overlapping halos from underwater light sources, which hindered the YOLOv10 model's ability to accurately determine target locations without NMS, ultimately affecting the accuracy of centroid positioning.

Figure 9 illustrates the detection performance of various methods across different scenarios. In scenario 1, where light sources form non-circular spots due to viewing angles and narrow beam spread, deep learning methods can identify the deformed light sources, while traditional image processing methods incorrectly filter them out. Furthermore, our algorithm achieves a target light source confidence level of 0.85, significantly higher than other deep learning algorithms. It also accurately locates the optical center, whereas YOLOv8 with the centroid method often misidentifies the optical center due to reliance on superficial brightness or shape features. When light source deformation is minimal and sources are mostly circular, as in scenario 2, all methods detect the light sources with similar accuracy. However, the simplistic network structure of CNN leads to instability in optical center detection. Traditional image processing methods and YOLOv8 with

centroid method tend to approximate the optical center as the geometric center. In contrast, our algorithm leverages high-dimensional features from the surrounding area to assist in optical center localization, achieving better precision. In scenario 3, where reflections on the water surface are present, deep learning methods effectively distinguish between genuine and false light sources due to their advanced feature learning capabilities, whereas traditional image processing methods and CNN often misidentify these reflections as light sources. Overall, these results demonstrate the advantages of our algorithm in reducing target misidentifications and improving the accuracy of optical center detection.

Table 3. Results of comparative experiments.

Methods	AP <sub>iou50</sub>	AP <sub>iou50-95</sub>	AP <sub>oks50</sub>	FPS
Tradition	-	-	0.837	28
CNN	-	-	0.867	67.9
YOLOv8n + Centroid	0.911	0.572	0.881	40
YOLOv9t + Centroid	0.918	0.569	0.893	38.9
YOLOv10n + Centroid	0.924	0.555	0.871	51
YOLOv8n-pose	0.903	0.564	0.874	66.6
Ous	0.943	0.599	0.939	36.5

![](_page_12_Picture_4.jpeg)

Figure 9. Detection results of different methods. Each row from top to bottom corresponds to scenario 1, scenario 2, and scenario 3, respectively. (a) Ours; (b) YOLOv8n-pose; (c) YOLOv8n with centroid; (d) Tradition; (e) CNN.

#### 3.3.3. Ablation Experiments

Ablation studies are a common research methodology designed to elucidate the significance and function of individual components or features within a deep neural network model. To assess the effectiveness of various modules, we conducted ablation experiments on an underwater optical beacon dataset, with results presented in Table 4. Notably, the AP of keypoints was calculated concurrently with object box detection during training, ensuring that it was not influenced by the accuracy of object box detection, which may result in higher keypoint precision than box precision.

From Table 4, the YOLOv8n-pose baseline model achieves an AP<sub>iou50</sub> of 90.3% and an AP<sub>oks50</sub> of 96.6% on the underwater optical beacon dataset, demonstrating reliable performance in its default configuration. The introduction of a small target detection head resulted in a notable improvement in detecting smaller objects, particularly  $AP_{iou50}$ , from 90.3% to 92.3%. However, this enhancement was accompanied by an increase in computational complexity, with FLOP rising to 12.4 G and FPS decreasing to 49.6, demonstrating a trade-off between accuracy and inference speed. Further incorporation of dynamic

convolution and RFApose heads increased the AP<sub>iou50</sub> and AP<sub>iou50-95</sub> by 1.2%, while FLOP decreased to 10.5 G, but the increase in parameter further reduced the FPS to 38.6. Finally, introducing the keypoint joint IoU matching post-processing into the model further enhanced detection precision, with AP<sub>iou50</sub> reaching 94.3%, a 4% improvement over the baseline model, and keypoint detection AP<sub>oks50</sub> reaching 99.4%, a 2.8% increase. Despite the increase in FLOP and parameters, the model still operates at 36.5 FPS, meeting real-time performance requirements. This demonstrates that our algorithm maintains high detection accuracy while balancing computational resource utilization and inference speed well.

Table 4. Results of ablation experiments.

Model	AP <sub>iou50</sub>	AP <sub>iou50-95</sub>	AP <sub>oks50</sub>	FLOP (G)	Parameter (M)	FPS
YOLOv8n-pose	0.903	0.564	0.966	8.7	6.2	66.6
+p2 *	0.923	0.576	0.978	12.4	6.2	49.6
+p2, +DC *	0.928	0.581	0.978	11.8	7.5	43.7
+p2, +DC, +RFApose *	0.935	0.588	0.982	10.5	9.1	38.6
+p2, +DC, +RFApose, +kp *	0.943	0.599	0.994	10.5	9.1	36.5

\* +p2: adding a small object detection head at the P2 feature layer; +DC: replacing the standard convolutions with dynamic convolutions; +RFApoes: adding RFAConv to the pose detection head; +kp: using a NMS method based on keypoint joint iou matching.

# 4. Pose Estimation Based on Underwater Omnidirectional Rotating Optical Beacon

## 4.1. Azimuth Estimation

An underwater omnidirectional rotating optical beacon utilizing a rotational synchronized scanning method assists AUVs in determining their orientation relative to a docking station. The specific steps are as follows:

• By determining the frame rate of the AUV's camera, the observable angle of the optical beacon, and the permissible error margin, the maximum scanning rate of the omnidirectional rotating optical beacon can be calculated. The maximum scanning rate, *s*', is computed using the following equation:

$$s' = \frac{f \times a}{360^{\circ}} \tag{3}$$

where *f* represents the frame rate of the AUV's camera, and *a* is the observable angle of the beacon. In this work, the AUV's camera captures at 30 Hz, and the beacon's observable angle is 10°. Consequently, the maximum scanning rate that ensures the beacon's visibility by the AUV's camera at all times is s' = 0.83 rps. This means that when the scanning rate is less than 0.83 rps, the AUV's 30 Hz camera can always capture the output light at one angle position per complete rotation. Observational errors are inevitable due to factors like the camera's frame rate and light scatter. Thus, it is necessary to discuss the theoretical and actual angular positions of the beacon light's output due to these observational discrepancies. If the AUV's camera is within the allowed observational error range ( $\pm e$  degrees) of the beacon's central axis, the beacon is still considered to be facing the AUV directly. Figure 10 illustrates these error scenarios, and the formula for calculating the maximum positional error e' is as follows.

$$e' = \frac{360^\circ \times s}{f} + e \tag{4}$$

The precision of this formula is dependent on the scanning rate s and the permissible observational error e. When *e* is  $2^{\circ}$  and s is 0.5 rps, the maximum positional error is  $8^{\circ}$ , which is within acceptable limits. Therefore, the beacon's scanning rate is set to 0.5 rps in this work.

• During the docking process, due to the rotational characteristic of the beacon, the deep learning algorithm may detect the target light source in multiple consecutive frames within the same rotation. To accurately determine the beacon's orientation,

we propose a metric method based on the light source's characteristics, as detailed in Algorithm 2. We hypothesize that the larger the area of the detected target light source and the closer its shape to a circle, the more likely it is that the beacon is facing the AUV directly. Here, I denotes the input image, bboxarea represents the area of the bounding box, and bboxshape represents the aspect ratio of the bounding box. We assign weights to the metrics warea and wshape, and use their weighted sum as the final detection evaluation metric. Additionally, for occasional frame drops in continuous detection, interpolation is used to fill in the gaps.

![](_page_14_Figure_2.jpeg)

Figure 10. Error diagram.

Algorithm 2: Forward Beacon Detection Algorithm
<b>Input:</b> <i>I</i> , <i>bbox</i> , <i>w</i> <sub>area</sub> , <i>w</i> <sub>shape</sub> , <i>t</i>
<b>Output:</b> <i>bbox<sub>max</sub></i> , <i>t<sub>max</sub></i>
1 Initialization
2 <i>bbox, keypoint, t</i> $\leftarrow$ YOLOv8-UL( <i>I</i> )
3 detect, frame <sub>miss</sub> , frame <sub>det</sub> , {bbox} $\leftarrow$ False, 0, 0, []
4 if bbox is not none:
5 <i>detect, frame</i> <sub>miss</sub> $\leftarrow$ True, 0
$6  \{bbox\} \leftarrow \{bbox\} \cup bbox$
7 $frame_{det} = frame_{det} + 1$
8 <b>if</b> $frame_{det} \geq 5$ :
9 $\{bbox_{area}\} \leftarrow area(\{bbox\})$
10 $\{bbox_{shape}\} \leftarrow shape(\{bbox\})$
11 $score \leftarrow w_{area} * \{bbox_{area}\} + w_{shape} * \{bbox_{shape}\}$
12 <b>return</b> $bbox_{max}, t_{max} \leftarrow max$ ({ $bbox$ }, key = score)
13 <b>else:</b> $frame_{miss} = frame_{miss} + 1$
14 if $frame_{miss} \ge 5$ then
15 { <i>bbox</i> }, <i>detect</i> $\leftarrow$ [], False
16 end

• Prior to the start of AUV docking, the time synchronization between the AUV and the docking station is confirmed through a timing system. Then, using the detection time tdet of the beacon obtained in step 2 and the initial time tinit, the theoretical angular position of the beacon at any given moment can be calculated, representing the AUV's azimuth relative to the docking station. The formula is provided below:

$$A = ((t_{det} - t_{init}) \text{mod}s^{-1}) \times s \times 360^{\circ}$$
(5)

#### 4.2. Pose Estimation

The scanning method of the omnidirectional rotating optical beacon can determine the AUV's azimuthal orientation relative to the docking station. However, this method does not ascertain the precise pose of the AUV. This limitation arises because the broad field of view of the underwater camera allows the AUV to capture the optical beacon from various poses at the same location. To achieve accurate pose estimation, we employ the LOS method [32] to compute the AUV's pose. This method enables the calculation of equivalent angular deviations using a single light source, thereby deriving the AUV's current pose relative to the base station in both the horizontal and vertical planes.

$$\begin{cases} \alpha = \arctan\left(\frac{2u}{M} \times \tan \alpha_0\right) \\ \beta = \arctan\left(\frac{2v}{N} \times \tan \beta_0\right) \end{cases}$$
(6)

Here, (u,v) represents the pixel position of the optical center in the image, M is the number of rows in the image, and N is the number of columns. The symbols  $a_0$  and  $\beta_0$  denote the camera's field of view angles on the horizontal and vertical planes, respectively. The pixel position of the optical center can be directly obtained from the previously described detection algorithm. Based on the horizontal and vertical deviation angles, the AUV adjusts its heading to align with the target, typically by controlling its rudder or pitch. When the AUV is oriented toward the front of the docking station, it can directly adjust based on the deviation angles, continuously correcting its course to approach the station gradually. When the AUV is at the rear of the docking station, it must first turn to face the station before using the equivalent angular deviations for pose adjustments, ultimately achieving precise docking.

#### 5. Pool Experiments

To provide data support for the underwater omnidirectional rotating optical beacon docking system, this study designed and conducted a series of pool simulation experiments. These experiments aimed to evaluate the effectiveness of the proposed underwater omnidirectional rotating optical beacon docking method, particularly in terms of azimuth and relative pose estimation accuracy.

#### 5.1. Experiment Setup and Procedure

The experiments were conducted in a water pool measuring 80 m  $\times$  15 m  $\times$  30 m, characterized by relatively clear water with an attenuation coefficient of 0.40 m<sup>-1</sup>, which meets the standards for coastal seawater as referenced in [33]. To determine the specific orientation of the underwater optical beacon relative to the camera, an experimental measurement platform was designed and installed on a gantry above the pool. This platform was equipped with a rail system that allowed movement along the X-axis, facilitating the free translation of the camera in both the X and Y directions. Scales marked on the gantry and rails enabled experimenters to accurately record the camera's position on the XOY plane. For ease of measurement, the rotating beacon was placed on a small underwater platform 3 m below the water surface, ensuring the camera and rotating beacon were at the same depth. The experiments utilized an underwater camera, the HDMultiSeaCam (manufactured by DeepSea, based in San Diego, CA, USA), which has a frame capture rate of 30 Hz and a resolution of  $1920 \times 1080$ . The horizontal and vertical field of view angles are 85° and 50°, respectively, enabling effective capture of underwater target light sources. Figure 11 provides a top view of the experimental conditions for image acquisition. In addition, to explore the impact of natural light on the detection of underwater optical beacons, experiments were conducted during the day (with daylight) and at night (without daylight, with minimal artificial light). The daytime experiments were conducted between 1 PM and 3 PM when natural sunlight illuminated the water surface. The nighttime experiments took place between 8 PM and 9 PM, during which there was no sunlight, although

![](_page_16_Figure_2.jpeg)

a small amount of illumination was provided to ensure the experiments could proceed. The experimental procedure is as follows:

Figure 11. Experimental setup.

- Camera platform setup: The camera platform was mounted on the gantry above the pool, with the gantry's center serving as the origin for the X-axis. The camera moves along the X-axis from -4.5 m to 4.5 m in 3 m increments, moving four times and capturing video at each position to simulate the AUV viewing the optical beacon from different directions.
- Y-axis movement: With the position of the omnidirectional rotating optical beacon serving as the origin for the Y-axis, the camera platform moves along the Y-axis from 20 m to 50 m using the gantry, with 5 m increments, repeating the operation in step 1 seven times. This results in 24 video scans of the optical beacon from various positions. Figure 12 shows representative images collected during the experiment.

![](_page_16_Figure_7.jpeg)

**Figure 12.** Detection results of different methods. (a) Daylight, the beacon faces forward; (b) darkness, the beacon faces forward; (c) daylight, the beacon faces sideways; (d) darkness, the beacon faces sideways.

 Data collection: Experimental data were collected from offline videos with a resolution of 1920 × 1080. The algorithm runs on a vision computing board, Jetson AGX Orin (manufactured by NVIDIA, based in Santa Clara, CA, USA), which is compact, highly capable, and easily deployable in underwater robots. It provides efficient and reliable computational support for data processing and analysis.

#### 5.2. Experiment Results

The experimental results indicate that our underwater omnidirectional rotating optical beacon demonstrates excellent imaging capabilities at depths of up to 45 m, regardless of the presence of ambient light. As shown in Figure 12, under the same distance conditions, the imaging area of the light source is larger and its features are more pronounced in the absence of ambient light. At a detection distance of 50 m, the light source becomes nearly invisible in the presence of ambient light, while in dark conditions, the light source can still be detected within a certain angular range. This highlights the significant impact of lighting conditions on the effective range of the beacon, suggesting that it is more suitable for underwater environments with limited ambient light, such as seabed docking stations. Furthermore, the experiments were conducted in a clear water pool, minimizing the effects of light absorption and scattering caused by underwater particles. However, when the beacon is applied in turbid or algae-rich aquatic environments, its effective range may be further reduced.

The azimuthal errors between the actual and measured values of the underwater omnidirectional rotating optical beacon docking system under both daytime and nighttime lighting conditions are presented in Table 5.

X-Axis\Y-Axis	20 m	25 m	30 m	35 m	40 m	45 m	50 m
4.5m (day)	6.12	5.63	5.43	5.60	4.94	5.61	-
4.5m (night)	6.12	5.62	5.43	5.55	4.89	5.54	-
1.5m (day)	4.53	4.48	4.58	4.29	4.49	4.38	-
1.5m (night)	4.53	4.48	4.58	4.29	4.47	4.34	4.21
—1.5m (day)	3.92	3.73	3.86	3.15	2.95	2.78	-
-1.5m (night)	3.92	3.70	3.86	3.14	2.95	2.77	2.68
-4.5m (day)	5.85	5.52	5.55	5.43	5.17	5.58	-
-4.5m (night)	5.85	5.52	5.52	5.43	5.11	5.50	-

Table 5. Azimuth detection errors.

According to Table 5, the differences under varying lighting conditions are minimal, remaining within 0.1 degrees, and consistent trends are observed. However, when the distance increases to 40 m, detection accuracy at night is slightly higher than during the day, even with larger angular deviations (±4.5 degrees). This improvement is attributed to the enhanced imaging effect of the light source in the absence of daylight, which facilitates light source detection. At a distance of 50 m, the light source becomes nearly undetectable during the day, whereas it remains detectable at night with smaller angular deviations  $(\pm 1.5 \text{ degrees})$ . The table also indicates that errors at a distance of 20 m are larger compared to other distances, which might be attributed to the proximity affecting the image area of the light in the camera view. The area differences between different rotational angles are similar, leading to errors in beacon orientation determination. Additionally, errors at 4.5 m on the X-axis are generally higher than those at 1.5 m by approximately 2.5 degrees. This could be due to the smaller angle between the camera and the optical beacon at 1.5 m, which increases the likelihood of capturing the frontal light source compared to at 4.5 m. Overall, the average absolute error in azimuth is 4.54 degrees, with the maximum error at 6.12 degrees and the minimum error reaching 2.77 degrees. These results validate the effectiveness of the location information provided by the proposed omnidirectional rotating optical beacon and its accompanying scanning method. To assess the accuracy of the pose estimation algorithm based on the detection algorithm, Table 6 displays the errors

between actual and measured values of the equivalent horizontal deviation angles for the underwater omnidirectional rotating optical beacon docking system. Since the camera and the rotating beacon were maintained at the same depth during experiments, the equivalent vertical deviation angle was essentially zero and thus not recorded.

X-Axis\Y-Axis	20 m	25 m	30 m	35 m	40 m	45 m	50 m
4.5m (day)	4.26	4.13	4.01	3.88	3.95	3.93	-
4.5m (night)	4.26	4.13	4.01	3.82	3.90	3.87	-
1.5m (day)	2.15	2.91	1.97	2.28	1.72	1.95	-
1.5m (night)	2.15	2.91	1.97	2.28	1.72	1.94	1.84
—1.5m (day)	2.82	3.20	1.95	1.93	1.86	1.88	-
-1.5m (night)	2.81	3.20	1.95	1.93	1.85	1.84	1.79
—4.5m (day)	4.22	4.14	3.93	3.55	3.71	3.99	-
-4.5m (night)	4.22	4.14	3.90	3.54	3.62	3.95	-

Table 6. Equivalent horizontal deviation angle errors.

Based on the data presented in Table 6, the differences in error under various lighting conditions are negligible at close range. Nevertheless, as the distance increases and the imaging position approaches the edge of the image, the nighttime error is approximately 0.05 degrees smaller than that observed during the day. This reduction can be attributed to the clearer light path of the underwater light source at night, which facilitates more accurate centroid localization. We observe a general trend of decreasing error with increasing distance, which can likely be attributed to the reduction in the imaging area of the optical beacon. This reduction narrows the range of error in centroid detection and allows the imaging position to gradually align closer to the image center, thereby enhancing precision. In contrast, errors at a 4.5 m distance along the X-axis are consistently higher than those at 1.5 m. This discrepancy may be due to wide-angle distortion caused by the underwater camera's field of view. The calculations for equivalent deviation angles do not fully account for this distortion, resulting in reduced accuracy for off-center light source images. Overall, the average absolute error of the equivalent horizontal deviation angle is 3.09 degrees, with a maximum error of 4.26 degrees and a minimum error reaching 1.72 degrees. The average error of 3.09 degrees validates the high precision of our deep learning-based centroid localization algorithm, meeting the requirements for engineering experiments.

## 6. Conclusions

This study proposes a deep learning-based approach for AUV docking using an omnidirectional rotating optical beacon, designed to address issues related to limited emission angles and insufficient beacon detection accuracy encountered during underwater optical docking. By employing an omnidirectional rotating optical beacon, the system achieves comprehensive 360-degree light radiation and significantly increases the likelihood of beacon detection by AUVs using a scanning method, thereby providing accurate azimuth information and two degrees of freedom. Additionally, this paper introduces a parallel deep-learning detection algorithm for optical beacons and centroids enhanced by the YOLOv8-pose model. This approach significantly improves the multi-scale positioning accuracy of optical beacons through optimized network structures and advanced post-processing techniques, ensuring stability and robustness in complex underwater environments. Results from pool experiments indicate that the designed system can provide high-precision omnidirectional docking within a range of 45 m, with the proposed algorithm outperforming traditional methods in terms of detection accuracy and processing speed. Compared to baseline models, our detection strategy enhances target detection and keypoint localization accuracy by 4% and 2.8%, respectively, while substantially reducing the false detection rate. Moreover, a motion trend measurement method based on the characteristics of the light source has been implemented, accurately detecting the orientation of the rotating optical beacon. Overall, the docking scheme performs as expected, offering a robust and efficient solution for omnidirectional autonomous AUV docking.

In future work, we intend to incorporate a broader variety of optical beacons and datasets from complex environments to enhance the adaptability and generalization capability of the algorithm. Additionally, we will improve the design of the omnidirectional rotating optical beacon to facilitate both horizontal 360-degree scanning and vertical scanning, thereby expanding its guidance range. System validation and optimization will be conducted using AUVs in more challenging real-world marine environments to ensure the practicality of this technology. To address potential adaptability in various underwater environments, we will explore the integration of adaptive light source adjustment technology and underwater image enhancement algorithms. These measures aim to improve interference resistance in turbid conditions by optimizing light source modulation and detection strategies, ensuring reliable localization under complex scenarios. Ultimately, our research will focus on further enhancing the stability and reliability of the docking scheme for the omnidirectional rotating optical beacon, advancing the development of autonomous AUV docking technologies, and providing robust support for AUV missions that require multi-angle and long-distance guidance and localization, such as marine monitoring, rescue operations, and deep-sea sampling.

**Author Contributions:** Conceptualization, Y.L. and K.S.; methodology, Y.L.; software, Y.L.; formal analysis and investigation, Y.L., K.S., Z.H. and J.L.; writing—original draft preparation, Y.L.; writing—review and editing, K.S. and Z.H.; validation, Y.L. and J.L.; project administration and funding acquisition, K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data will not be made available.

**Acknowledgments:** The authors gratefully acknowledge the support extended by the State Key Laboratory of Robotics, Shenyang, in encouraging this research work.

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

#### References

- Wynn, R.B.; Huvenne, V.A.; Le Bas, T.P.; Murton, B.J.; Connelly, D.P.; Bett, B.J.; Ruhl, H.A.; Morris, K.J.; Peakall, J.; Parsons, D.R.; et al. Autonomous Underwater Vehicles (AUVs): Their past, present and future contributions to the advancement of marine geoscience. *Mar. Geol.* 2014, 352, 451–468. [CrossRef]
- 2. Sun, K.; Cui, W.; Chen, C. Review of underwater sensing technologies and applications. Sensors 2021, 21, 7849. [CrossRef]
- 3. Hou, Y.; Han, G.; Zhang, F.; Lin, C.; Peng, J.; Liu, L. Distributional Soft Actor-Critic-Based Multi-AUV Cooperative Pursuit for Maritime Security Protection. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 6049–6060. [CrossRef]
- Li, Y.; Sun, K. Review of underwater visual navigation and docking: Advances and challenges. In Proceedings of the Sixth Conference on Frontiers in Optical Imaging and Technology, Nanjing, China, 22–24 October 2023; SPIE: Bellingham, WA, USA, 2024; Volume 13156, pp. 314–321.
- Yan, Z.; Gong, P.; Zhang, W.; Li, Z.; Teng, Y. Autonomous Underwater Vehicle Vision Guided Docking Experiments Based on L-Shaped Light Array. *IEEE Access* 2019, 7, 72567–72576. [CrossRef]
- 6. Zhang, W.; Wu, W.; Teng, Y.; Li, Z.; Yan, Z. An underwater docking system based on UUV and recovery mother ship: Design and experiment. *Ocean Eng.* **2023**, *281*, 114767. [CrossRef]
- Trslic, P.; Rossi, M.; Robinson, L.; O'Donnel, C.W.; Weir, A.; Coleman, J.; Toal, D. Vision based autonomous docking for work class ROVs. Ocean Eng. 2020, 196, 106840. [CrossRef]
- Cheng, H.; Chu, J.; Zhang, R.; Gui, X.; Tian, L. Real-time position and attitude estimation for homing and docking of an autonomous underwater vehicle based on bionic polarized optical guidance. J. Ocean Univ. China 2020, 19, 1042–1050. [CrossRef]
- Zhao, C.; Dong, H.; Wang, J.; Qiao, T.; Yu, J.; Ren, J. Dual-Type Marker Fusion-Based Underwater Visual Localization for Autonomous Docking. *IEEE Trans. Instrum. Meas.* 2024, 73, 1–11. [CrossRef]
- 10. Sun, K.; Han, Z. Autonomous underwater vehicle docking system for energy and data transmission in cabled ocean observatory networks. *Front. Energy Res.* 2022, 10, 960278. [CrossRef]
- 11. Lin, M.; Lin, R.; Yang, C.; Li, D.; Zhang, Z.; Zhao, Y.; Ding, W. Docking to an underwater suspended charging station: Systematic design and experimental tests. *Ocean Eng.* 2022, 249, 110766. [CrossRef]
- 12. Zhang, Z.; Ding, W.; Wu, R.; Lin, M.; Li, D.; Lin, R. Autonomous Underwater Vehicle Cruise Positioning and Docking Guidance Scheme. J. Mar. Sci. Eng. 2024, 12, 1023. [CrossRef]

- Cai, C.; Rong, Z.; Xie, X.; Xu, B.; Zhang, Z.; Wu, Z.; Si, Y.; Huang, H. Development and test of a subsea docking system applied to an autonomous underwater helicopter. In Proceedings of the OCEANS 2022, Hampton Roads, VA, USA, 17–20 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–7.
- Dörner, D.; Espinoza, A.T.; Torroba, I.; Kuttenkeuler, J.; Stenius, I. To Smooth or to Filter: A Comparative Study of State Estimation Approaches for Vision-Based Autonomous Underwater Docking. In Proceedings of the OCEANS 2024, Singapore, 15–18 April 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–9.
- 15. Xu, S.; Jiang, Y.; Li, Y.; Wang, B.; Xie, T.; Li, S.; Cao, J. A stereo visual navigation method for docking autonomous underwater vehicles. *J. Field Robot.* **2024**, *41*, 374–395. [CrossRef]
- Wang, S.; Wang, X.; Lei, P.; Chen, J.; Xu, Z.; Yang, Y.; Zhou, Y. Blue laser diode light for underwater optical vision guidance in AUV docking. In Proceedings of the Semiconductor Lasers and Applications IX, Hangzhou, China, 20–23 October 2019; SPIE: Bellingham, WA, USA, 2019; Volume 11182, pp. 175–183.
- 17. Zhang, Y.; Wang, X.; Lei, P.; Wang, S.; Yang, Y.; Sun, L.; Zhou, Y. Smart vector-inspired optical vision guiding method for autonomous underwater vehicle docking and formation. *Opt. Lett.* **2022**, *47*, 2919–2922. [CrossRef] [PubMed]
- Chen, Y.; Duan, Z.; Zheng, F.; Guo, Y.; Xia, Q. Underwater optical guiding and communication solution for the AUV and seafloor node. *Appl. Opt.* 2022, *61*, 7059–7070. [CrossRef] [PubMed]
- 19. Lv, F.; Xu, H.; Shi, K.; Wang, X. Estimation of Positions and Poses of Autonomous Underwater Vehicle Relative to Docking Station Based on Adaptive Extraction of Visual Guidance Features. *Machines* **2022**, *10*, 571. [CrossRef]
- Feng, J.; Yao, Y.; Wang, H.; Jin, H. Multi-AUV Terminal Guidance Method Based on Underwater Visual Positioning. In Proceedings of the 2020 IEEE International Conference on Mechatronics and Automation (ICMA), Beijing, China, 13–16 October 2020; pp. 314–319.
- Li, Y.; Jiang, Y.; Cao, J.; Wang, B.; Li, Y. AUV Docking Experiments Based on Vision Positioning Using Two Cameras. *Ocean Eng.* 2015, 110, 163–173. [CrossRef]
- 22. Zhang, B.; Zhong, P.; Yang, F.; Zhou, T.; Shen, L. Fast Underwater Optical Beacon Finding and High Accuracy Visual Ranging Method Based on Deep Learning. *Sensors* 2022, 22, 7940. [CrossRef]
- 23. Ren, R.; Zhang, L.; Liu, L.; Yuan, Y. Two AUVs Guidance Method for Self-Reconfiguration Mission Based on Monocular Vision. *IEEE Sensors J.* 2021, *21*, 10082–10090. [CrossRef]
- Chavez-Galaviz, J.; Mahmoudian, N. Underwater Dock Detection Through Convolutional Neural Networks Trained with Artificial Image Generation. In Proceedings of the 2022 International Conference on Robotics and Automation, Philadelphia, PA, USA, 23–27 May 2022; pp. 4621–4627.
- 25. Duntley, S.Q. Light in the Sea. J. Opt. Soc. Am. 1963, 53, 214–233. [CrossRef]
- Liu, S.; Ozay, M.; Okatani, T.; Xu, H.; Sun, K.; Lin, Y. Detection and Pose Estimation for Short-Range Vision-Based Underwater Docking. *IEEE Access* 2018, 7, 2720–2749. [CrossRef]
- Chen, Y.; Dai, X.; Liu, M.; Chen, D.; Yuan, L.; Liu, Z. Dynamic Convolution: Attention Over Convolution Kernels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11030–11039.
- 28. Zhang, X.; Liu, C.; Yang, D.; Song, T.; Ye, Y.; Li, K.; Song, Y. RFAConv: Innovating Spatial Attention and Standard Convolutional Operation. *arXiv* 2023, arXiv:2304.03198.
- 29. Zhang, M.; Xu, S.; Song, W.; He, Q.; Wei, Q. Lightweight Underwater Object Detection Based on YOLOv4 and Multi-Scale Attentional Feature Fusion. *Remote Sens.* **2021**, *13*, 4706. [CrossRef]
- Neubeck, A.; Van Gool, L. Efficient Non-Maximum Suppression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; Volume 3, pp. 850–855.
- Maji, D.; Nagori, S.; Mathew, M.; Poddar, D. YOLO-Pose: Enhancing YOLO for Multi-Person Pose Estimation Using Object Keypoint Similarity Loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LO, USA, 18–24 June 2022; pp. 2637–2646.
- Su, X.; Xiang, X.; Dong, D.; Zhang, J. Visual LOS Guided Docking of Over-Actuated Underwater Vehicle. In Proceedings of the Global Oceans 2020: Singapore–US Gulf Coast, Biloxi, MS, USA, 5–30 October 2020; pp. 1–5.
- 33. Petzold, T.J. Volume Scattering Functions for Selected Ocean Waters; Scripps Institution of Oceanography: San Diego, CA, USA, 1972.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.