

A study of the landforms and megafaunal characteristics of the Caiwei Guyot area by manned submersible image data-driven technology

Zhongjun Ding^{1, 2*}, Xingyu Wang^{1, 2}, Chen Liu^{1, 2}, Guangyang Ma^{1, 2}, Chanjuan Cao^{1, 2}

¹ National Deep Sea Center, Qingdao 266237, China

² College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Received 18 June 2024; accepted 1 December 2024

© Chinese Society for Oceanography and Springer-Verlag GmbH Germany, part of Springer Nature 2025

Abstract

Scientific and precise evaluations of the megafaunal and landform characteristics of seamounts are important guides for their protection and study. A series of manned and unmanned submersibles have provided invaluable observational imaging data for the ecological study of seamounts. However, traditional methods of artificial observation of seamount imaging data cannot accurately and efficiently determine the characteristics of megafauna and landforms. This research harnesses data-driven technology to systematically investigate the distributional traits and morphological features of megafaunal organisms, as well as the topographical characteristics, in the Caiwei Guyot region of the western Pacific's Magellan Seamounts. To construct the landform and megafauna dataset of the Caiwei Guyot region, we used a data preprocessing technology based on image enhancement to provide high-quality imaging data for data-driven technologies. A megafaunal identification and counting algorithm based on YOLOv5 (You Only Look Once Version 5) was developed to efficiently assess the abundance, variety, and dominant species of megafauna. Simultaneously, a landform three-dimensional (3D) reconstruction algorithm based on PatchmatchNet was developed to reconstruct the 3D form of the terrain accurately. This study pioneers the application of data-driven technology to deep-sea imaging within the Caiwei Guyot region, offering an innovative approach to accurately and efficiently characterize the region's unique megafauna and landforms.

Key words manned submersible imaging, data-driven, Caiwei Guyot, landforms, megafauna

Citation Ding Zhongjun, Wang Xingyu, Liu Chen, Ma Guangyang, Cao Chanjuan. 2025. A study of the landforms and megafaunal characteristics of the Caiwei Guyot area by manned submersible image data-driven technology. *Acta Oceanologica Sinica*, 44(1): 147–164, doi: 10.1007/s13131-024-2399-3

1 Introduction

Seamounts are renowned as the “great garden under the sea” due to their rich mineral and biological resources. It is of great significance to study the biological and geomorphic characteristics of the seamount area for the protection and development of deep-sea resources (Ding, 2021; Li et al., 2018; Liu et al., 2018). However, traditional deep-sea geomorphology and biological survey methods primarily involve labor-intensive manual sampling and direct visual inspection of deep-sea sonar and optical images, which is inefficient and narrow in applicability.

Therefore, to improve the automation of data processing of deep-diving images and obtain more accurate analysis results, we must design an intuitive and effective data processing method.

Numerous academics have recently researched the attributes of marine life and landforms. Geyer (1977) generated a distribution map of deep-sea biological communities on the DeSoto Canyon-Louisiana Continental Slope using images obtained by the DIAPHUS manned submersible. Collot et al. (1992) utilized observational data collected by the manned submersible *Paralcyon* to construct a geological profile of the steep slope of the newly

Foundation item: The Key Research and Development Program of Shandong Province of China under contract No. 2020JMRH0101; the National Key Research and Development Project of China under contract No. 2021YFC2802100; the Qingdao Natural Science Foundation under contract No. 24-4-4-zrij-127-jch.

*Corresponding author, E-mail: dzj@ndsc.org.cn

<http://www.aosocean.com>
E-mail: ocean2@hyxb.org.cn

discovered Hecate Ridge in the North Fiji Basin, which provided valuable insight into diverse aspects, such as biological populations and fishery resources. Usui et al. (1993) conducted manned deep-sea diving observation and sampling research in the Tenpo Seamount in the western Pacific Ocean by utilizing Deep-sea 2000 Manned submersible. Along the submersible's navigation track, they investigated the distribution of manganese nodules and manganese-crusts coastal mountain slopes. This research enhanced our knowledge of metal mineral resources in seamounts. Duan et al. (2023) employed side-scan sonar techniques to investigate the micro-geomorphological characteristics of the seafloor within the survey area and obtained a vivid representation of the seabed features in a specific maritime zone south of Zhuhai, China. Their research provided a detailed visual dataset that enhanced our understanding of underwater topography and its implications for benthic ecology in the region. Zhang et al. (2021) obtained precise bathymetric landforms and shallow stratigraphic profile data in the survey area using high-resolution submarine acoustic detection equipment. They also determined the surface sediment type and particle size parameters using a sieve analysis method and a laser particle size meter experiment. Ottesen et al. (2017) investigated the seafloor morphology of nine tidewater glaciers terminating in the open sea area east of Svalbard based on multi-beam band-sounding technology and then studied the evolution of the glaciers.

With the development of underwater cameras, as well as the advantages of underwater optical detection methods that can capture details such as color information of targets, some scholars study biological and geomorphological features based on underwater optical videos or images (Xie et al., 2022). Usui et al. (2013) conducted a methodical investigation and sampling campaign on Marcus Seamount in the western Pacific Ocean. They utilized manned and remotely operated submersibles and deep-towed equipment to take optical video of the seafloor. They integrated their research findings with the submersible's observation track to analyze the entire distribution of crust resources, encompassing water depth, crust thickness, and sample properties. Liu et al. (2023) characterized the topographic and geomorphic features of the Caroline M4 Seamount and its surrounding area based on comprehensive coverage multibeam bathymetry data and optical imagery acquired through multiple descents of a remotely operated vehicle (ROV) in the offshore region. Baker et al. (2021) combined directional observations from manned submersibles with quantitative analysis of fixed images obtained from stereo cameras to estimate the abundance, distribution, habitat, and behavior of Pacific Ocean fish. Perez et al. (2018) analyzed video data recorded during their dives to characterize the assemblage of benthic megafauna (fish and crustaceans) and to quantitatively assess structural factors (depth, topography, and habitat type). Grinyó et al. (2022) studied the geomorphic structure and giant benthos of the Gulf of Cadiz with 17 videos of dives taken by an ROV. The distribution of the

deep-sea ecosystem of the Gulf of Cadiz was obtained.

The Caiwei Guyot, located in the northwest Pacific, is a deep flat-topped seamount with a summit area rich in cobalt-rich crusts, which hold significant mineral resource value. Moreover, due to its unique marine and hydrographic conditions, the Caiwei Guyot considerably influences the distribution of deep-sea ecosystems, biodiversity, and geological features on the seafloor. Studying the Caiwei Guyot enables scientists to understand how seamount topography affects local circulation patterns and how these patterns regulate the exchange of energy and materials within the ocean (Guo et al., 2020). However, there are currently few reports on the topography and large benthic organism characteristics of the Caiwei Guyot, which holds abundant research potential.

In addition, the extraction of information on deep-sea organisms and topographical features from manned submersible observation imagery has traditionally relied on manual analysis of offline video footage. While this approach benefits from integrating human prior knowledge to analyze complex scenes reliably, it is hampered by low information extraction efficiency, the inability to rapidly and intuitively obtain seabed information, and the challenges of achieving high-precision quantitative descriptions due to human subjectivity. In recent years, the evolution of data-driven methods and artificial intelligence technologies has led to intelligent techniques for efficiently processing underwater imagery, offering significant efficiency, precision, and cost advantages. These advancements are increasingly becoming the cutting-edge in deep-sea survey methodologies. Therefore, we integrate data-driven technologies to address the aforementioned shortcomings of manual observation. By mitigating errors caused by the subjectivity of investigators, we enhance the speed and accuracy of data processing. Taking the topographical and biological survey of the Caiwei Guyot as an example, we demonstrate the advantages of data-driven technology over traditional manual observation in deep-sea surveys. This aims to provide a novel technical reference for the efficient utilization of deep-sea survey data and the automated extraction of deep-sea information.

The structure of this paper is as follows. The second section provides a brief introduction to the sources of our experimental data, as well as the principles and procedures of the data-driven approach used in this study. The third section discusses the basic parameters of the experimental equipment employed, the information processing effects of the data-driven method, and the research findings on the distribution of biology and topography in the Caiwei Guyot facilitated by this method. Finally, the fourth part summarizes the content of this paper.

2 Materials and methods

2.1 Survey area and means

The Caiwei Guyot is located in the Northwest Pacific Ocean, east of the Mariana Island Arc, and approximately

1 100 km from Guam and 1 000 km from the port of Pohnpei in the south. The *Jiaolong* manned submersible used the unique technical advantages in the Caiwei Guyot area. According to the idea of “point, line, and surface”, combined with an anatomical investigation, typical terrain units, scientific selection of diving sites, and reasonable arrangement of operation content, the *Jiaolong* manned submersible has carried out a detailed investigation of the typical terrain units of the peaks, slopes, and foothills in the south, north, and west of the seamount. Eight dives were conducted, and the topographic map of the Caiwei Guyot and dive sites are shown in Fig. 1. Among them, the red dots represent eight distinct dive sites.

The *Jiaolong* manned submersible used six deep-sea high-definition cameras with a resolution of 1080i HDTV to shoot a video of the Caiwei Guyot area. The minimum illumination is 2 lx. The optical zoom is 10 $\times$. The digital zoom is 12 $\times$. The aperture range is 3.2–32 mm, and the video aspect ratio is 16:9 or 4:3. It has a maximum speed of 2.5 kn during shooting and a cruise speed of 1 kn, and it uses 10 LEDs and a quartz halogen lamp to provide lighting. During this process, a large amount of valuable deep-sea video data was obtained.

2.2 Methods

2.2.1 Overall architecture of the method

Data-driven technology is based on large-scale datasets that use machine and deep learning techniques to extract useful information from data and make analyses and predictions (Hou and Xu, 2009). Consequently, this article presents a data-driven methodology for analyzing deep-diving image data to examine the megafauna characteristics and landforms of the Caiwei Guyot region.

Figure 2 illustrates the overall architecture of the method, which primarily comprises four components: (1) megafaunal identification, (2) a counting algorithm based on YOLOv5, (3) a landform 3D reconstruction algorithm based on PatchmatchNet, and (4) analysis of the results.

The primary objective of the data preprocessing phase is to generate an image dataset from videos and enhance the image data's quality using the algorithm. This is done to supply the succeeding data-driven model with more distinct image data. The megafaunal identification and counting algorithm based on YOLOv5 detect the category and quantity of megafaunal organisms in the image data of the Caiwei Guyot region quickly and precisely. The landform 3D reconstruction algorithm based on PatchmatchNet performs multi-view 3D reconstruction of a typical landform rapidly and accurately. The analysis of the results relies on the outputs of the two data-driven models to determine the characteristics of the megafauna and landforms in the Caiwei Guyot.

The presence of light attenuation and dispersion in the deep sea, coupled with limited illumination, inevitably results in color bias and low contrast in deep-sea images. Degraded deep-sea images can negatively affect the efficacy of data-driven models. Therefore, we propose an image enhancement method to furnish data-driven models with high-quality deep-sea imagery. The procedure commences by applying the gray world method (Ancuti et al., 2018) to rectify the image's color bias. Subsequently, the color-corrected image is converted from RGB space to LAB space. The CLAHE algorithm is then utilized to augment the L channel. Ultimately, the deep-sea image is enhanced. Furthermore, due to the unique characteristic of

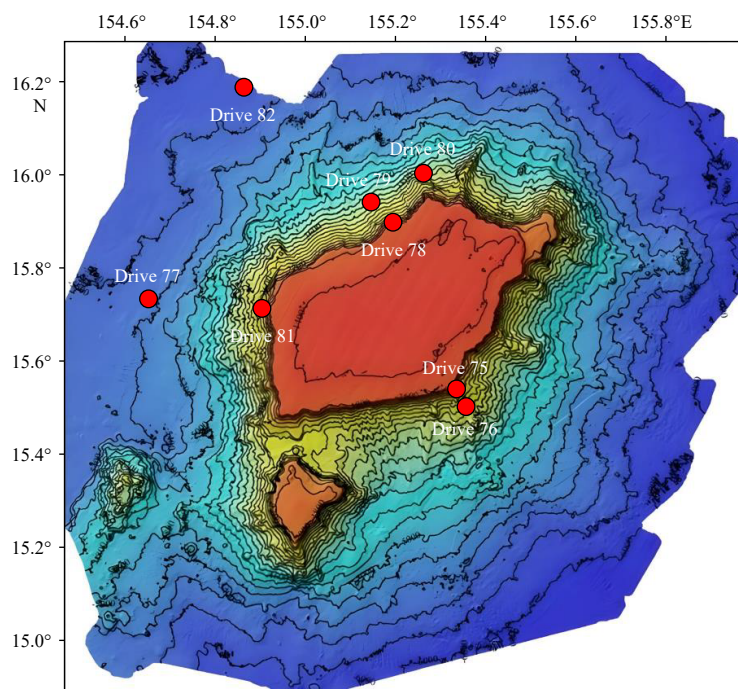


Fig. 1. Topographic map of the Caiwei Guyot.

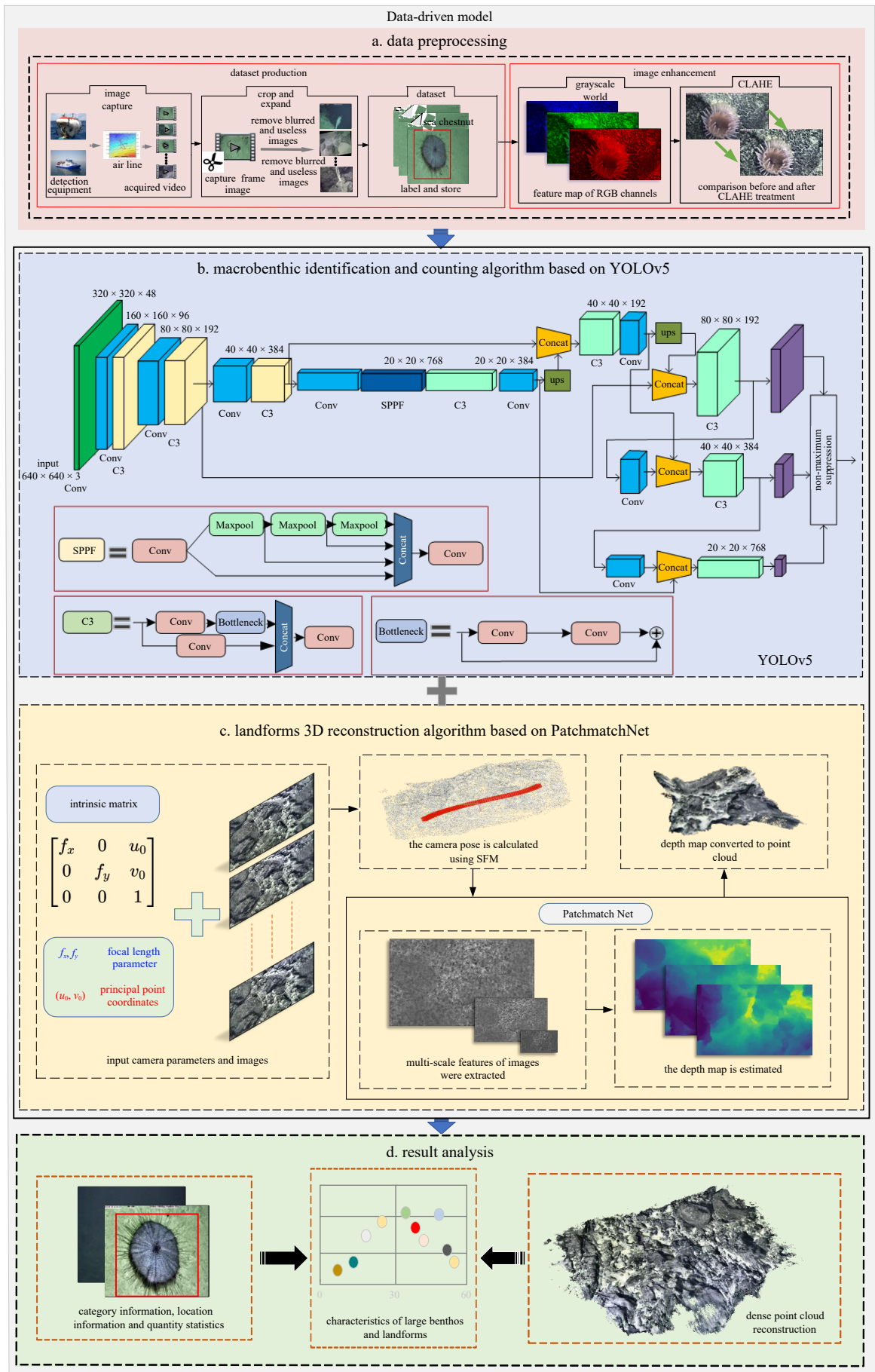


Fig. 2. Method architecture.

visible light propagation attenuation, the average value of pixels in the red channel of deep-sea images is comparatively low. Conversely, the grayscale world assumes that the average value of pixels across various channels in the scene should be equivalent. Thus, the red channel will be inflated to an inordinate degree when the grayscale world method is utilized to restore color to the deep-sea images. Consequently, color artifacts will appear in the enhanced images. To address this, the red channel is initially pre-compensated using Eq. (1):

$$I_{rc}(x) = I_r(x) + \alpha \cdot (\bar{I}_g - \bar{I}_r) \cdot (1 - I_r(x)) \cdot I_g(x), \quad (1)$$

where $I_{rc}(x)$ and $I_r(x)$ are the values of the red channel after compensation and before compensation, respectively, $\bar{I}_g$ and $\bar{I}_r$ are the mean values of the blue and red channels, respectively, and α indicates the constant parameter. After the red channel is pre-compensated, the gray world method is used to correct the color bias of the image. This method adjusts the color of each pixel by calculating the average gray value of the image (the average value of all pixel values) so that the gray value of each pixel corresponds to the average gray value of the entire image to achieve the color balance of the image. Specifically, we first calculate the average values of the RGB channels, denoted as $\bar{R}$, $\bar{G}$, and $\bar{B}$, to derive the white balance coefficient K . Subsequently, we determine the gain factors for the RGB channels, which are expressed as follows:

$$\begin{cases} \text{gain}_R = \frac{K}{\bar{R}}, \\ \text{gain}_B = \frac{K}{\bar{B}}, \\ \text{gain}_G = \frac{K}{\bar{G}}. \end{cases} \quad (2)$$

Finally, each pixel value can be adjusted according to the gain coefficient to achieve image color correction:

$$\begin{cases} R' = \text{gain}_R \times R, \\ B' = \text{gain}_B \times B, \\ G' = \text{gain}_G \times G. \end{cases} \quad (3)$$

The CLAHE algorithm (Wang and Niu, 2022) is implemented on deep-sea images to enhance their contrast and luminosity. By partitioning the image into small blocks and limiting the number of pixels in each gray level, CLAHE achieves a more uniform distribution of gray levels within each block. Interpolation is then used to produce an image with increased contrast. The L channel, which represents brightness information, is isolated from the A and B channels, which represent color information, during local contrast enhancement. As a result, CLAHE processes solely the L channel associated with contrast, thus potentially mitigating color distortion and computational complexity. The discrete stages comprising the CLAHE algorithm are as follows.

(1) Divide the picture into continuous and non-coinci-

dent $m \times n$ subblocks.

(2) Obtain histogram $h(x)$ of the subblocks using the direct square equalization method (Pizer et al., 1987).

(3) Calculate the clipping limit T :

$$T = C_{\text{clip}} \times \frac{N_x \times N_y}{M}, \quad (4)$$

where N_x and N_y are the number of pixels of each subblock in the x and y directions, respectively, M is the gray level of the corresponding subblock, and C_{clip} is the clipping coefficient.

(4) Cut the gray histogram of the subblocks according to T and divide the cut pixels into each gray level of the gray histogram according to Eq. (5) and Eq. (6):

$$S = \sum_{x=0}^{M-1} \{\max[h(x) - T]\}, \quad (5)$$

$$A = \frac{S}{M}, \quad (6)$$

where A is the pixels evenly assigned to each gray level, and S is the total number of pixels exceeding T . The histogram $H(x)$ after reallocation is:

$$H(x) = \begin{cases} T + A & h(x) \geq T, \\ h(x) + A & h(x) < T. \end{cases} \quad (7)$$

(5) Obtain the gray value of the center point of each subblock, use it as the anchor value, and reconstruct the remaining pixels in the image using the double-line interpolation method (Catmull, 1974).

2.2.2 Megafaunal identification and counting algorithm based on YOLOv5

The disadvantages of the method for classifying and quantifying megafauna using artificial observations of seamount image data include inefficiency and subjectivity in judgment. Object detection technology based on deep learning can precisely identify and localize targets by extracting target features from images. This can provide efficient and rapid data support for the study of megafauna (Ding et al., 2023). Consequently, we present the megafaunal identification and counting algorithm based on YOLOv5. This algorithm enables the classification and counting of megafaunal in image data with speed and precision and facilitates subsequent statistical analysis of biological characteristics. In particular, the video data obtained during each voyage segment of the *Jiaolong* manned submersible were utilized to construct an extensive database on deep-sea benthos. This database was subsequently employed to train the YOLOv5 model. In conclusion, the trained YOLOv5 model was applied to the image data of the Caiwei Guyot region to classify and quantify the megafauna.

The principle of the typical one-stage algorithm

YOLOv5 (Jocher, 2020) is illustrated in Fig. 3. Bounding boxes are predicted in each grid of the original input image to detect distinct categories of targets. The output comprises the bounding boxes corresponding to each target type, with the confidence of each bounding box being computed individually. The process involves establishing a threshold to exclude bounding boxes with class confidence below the specified value. Non-maximum suppression is then utilized to obtain the final bounding box. Thus, this model can be utilized to extract the category and quantity information on every organism depicted in each frame image.

YOLOv5 primarily comprises an input terminal, backbone network, neck network, and detection head. The structure is illustrated in Fig. 3. Image preprocessing methods, including the adaptive anchor frame calculation and mosaic data enhancement, were implemented at the input end. The backbone network, consisting of the Conv, C3, and spatial pyramid pooling fast (SPPF) structures, is utilized for feature extraction at various target dimensions. The feature pyramid network (FPN) and path aggregation network (PAN) featuring a fusion structure are implemented in the neck. FPN transmits high-level semantic features from the top to the bottom, enabling the bottom feature map to acquire semantic information from the top layer. PAN is then used to transmit features, such as position and texture, from the bottom to the top, allowing the top-level feature map to obtain additional position and texture information from the low-level feature map.

The detection head is the component responsible for mapping the extracted features into the target positional coordinates and category confidence. This structure optimizes the model parameters by computing the loss function, thus outputting information on the target's location, category, and confidence. The expression for the total loss function, L , is as follows:

$$L = l_{\text{conf}}(c_i^{\text{gt}}, \hat{c}_i) + l_{\text{cls}}(p_i^{\text{gt}}, \hat{p}_i) + l_{\text{coord}}(b_i^{\text{gt}}, \hat{b}_i), \quad (8)$$

where l_{conf} represents the confidence loss, l_{cls} represents

the classification loss, and l_{coord} is the bounding box regression loss. $\hat{c}_i$ and c_i^{gt} are the predicted and actual confidence levels of an object that is presented in the i -th grid cell, respectively. $\hat{p}_i(z)$ and $p_i^{\text{gt}}(z)$ denote the predicted and actual probabilities that the object in the i -th grid cell belongs to category z , respectively. $\hat{b}_i$ and b_i^{gt} correspond to the coordinates of the predicted and actual bounding boxes within the i -th grid cell, respectively. The expression for l_{conf} is as follows:

$$l_{\text{conf}}(c_i^{\text{gt}}, \hat{c}_i) = - \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [c_i^{\text{gt}j} \ln(\hat{c}_i^j) + (1 - c_i^{\text{gt}j}) \ln(1 - \hat{c}_i^j)] - \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} [c_i^{\text{gt}j} \ln(\hat{c}_i^j) + (1 - c_i^{\text{gt}j}) \ln(1 - \hat{c}_i^j)]. \quad (9)$$

In the equation, S^2 refers to the number of grids, B denotes the number of bounding boxes per grid cell, and I_{ij}^{obj} indicates whether the j -th bounding box in the i -th grid is responsible for detecting the object. $\hat{c}_i^j$ and $c_i^{\text{gt}j}$ represent the predicted and actual confidences, respectively, that an object is present within the j -th bounding box of the i -th grid. The expression for l_{cls} is as follows:

$$l_{\text{cls}}(p_i^{\text{gt}}, \hat{p}_i) = - \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \sum_{z \in \text{classes}} [p_i^{\text{gt}j}(z) \lg(\hat{p}_i^j(z)) + (1 - p_i^{\text{gt}j}(z)) \lg(1 - \hat{p}_i^j(z))]. \quad (10)$$

In the given context, $\hat{p}_i^j(z)$ and $p_i^{\text{gt}j}(z)$ correspond to the predicted and actual probabilities, respectively, that the object within the j -th bounding box of the i -th grid belongs to category z . The expression for l_{coord} is as follows:

$$l_{\text{coord}}(b_i^{\text{gt}}, \hat{b}_i) = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left(1 - \text{IoU} + \frac{\rho^2(j_i^{\text{gt}}, \hat{j}_i)}{d^2} + \frac{\frac{16}{\pi^3} (\arctan w^{\text{gt}}/h^{\text{gt}} - \arctan \hat{w}/\hat{h})^3}{(1 - \text{IoU}) + \frac{4}{\pi^2} (\arctan w^{\text{gt}}/h^{\text{gt}} - \arctan \hat{w}/\hat{h})^2} \right). \quad (11)$$

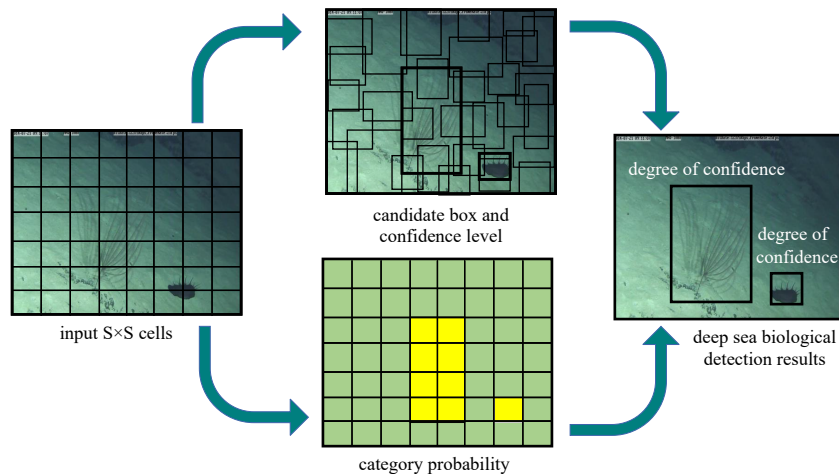


Fig. 3. Detection principle of YOLOv5.

In the equation, IoU represents the Intersection over Union (IoU) between the predicted and ground truth boxes. $w^{\text{gt}}/h^{\text{gt}}$ and $\hat{w}/\hat{h}$ denote the aspect ratios of the ground truth box and the predicted box, respectively. $\hat{j}_i$ and j_i^{gt} are the midpoint values of the predicted box and the ground truth box within the i -th grid, respectively. ρ^2 is the Euclidean distance between the two points. The term d refers to the diagonal distance of the smallest enclosing region that contains both the true and predicted boxes.

2.2.3 Landform 3D reconstruction algorithm based on PatchmatchNet

Understanding climate change, deep-sea landforms, and geomorphology is crucial to comprehending the marine ecosystem, biodiversity, and Earth's tectonic activity. In recent years, advancements in underwater photography technology have enabled deep-sea exploration to access a vast amount of imaging data. As a result, geographic information extraction via 3D visual reconstruction has become a hotspot of scientific inquiry.

Among them, multi-view stereo (MVS), based on its low cost, robustness, and scale adaptability, has been widely used in seabed landforms and geomorphology mapping. Traditional MVS methods typically need to input images with uniform illumination and robust texture features. In comparison, MVS methods based on deep learning are more robust to the above problems and have gradually attracted extensive attention from scholars internationally. Wang et al. (2021) learned from the idea of Patchmatch (Barnes et al., 2009) in the traditional MVS method and designed a learning-based Patchmatch architecture, PatchmatchNet, which considerably saves computing resources while maintaining reconstruction accuracy and has good comprehensive performance.

Imaging at great depths typically requires artificial light sources, which inevitably results in inconsistent illumination. Consequently, specific captured images will contain shadow regions that are challenging for image enhancement algorithms to rectify fully. Furthermore, the

restoration of 3D geomorphic information in a specific region necessitates processing many video images, and the memory demands of the employed methodology are critical. In light of these challenges, this approach implements a learn-based PatchmatchNet for multi-view 3D reconstruction to reconstruct the 3D structure of typical landforms rapidly and accurately.

The method is divided into four parts: camera pose estimation, depth map estimation, depth map fusion, and finally the reconstructed point cloud model. This study used the structure from motion (SFM) method to calculate the camera pose corresponding to the image as the input information of the PatchmatchNet depth map estimation. The precise procedure is illustrated in Fig. 4.

PatchmatchNet was utilized to estimate the depth of each image after obtaining the camera's pose information; the dense point cloud was subsequently reconstructed. The primary components of PatchmatchNet, as illustrated in Fig. 5, are the depth estimation module and the feature extraction module. To expedite the computation of depth maps, the feature extraction module employs a framework resembling the feature pyramid network to extract features in layers from multi-scale images. The depth estimation module generates the final depth estimation map after completing depth initialization, matching cost calculation and aggregation, depth map generation, and optimization.

To obtain the depth information of the current pixel more accurately, the network adopts the adaptive propagation method to propagate the assumed value of the depth of the pixel in the neighborhood, input the feature information of the reference image into the two-dimensional demorphable convolution network (Dai et al., 2017), learn the coordinates of the pixels in the same plane and the extra spatial offset of the current pixel, and estimate the depth selection range of the pixels to be processed in the next iteration:

$$D_p(p) = \{D(p + x_i + \Delta x_i(p))\}_{i=1}^{N_p}, \quad (12)$$

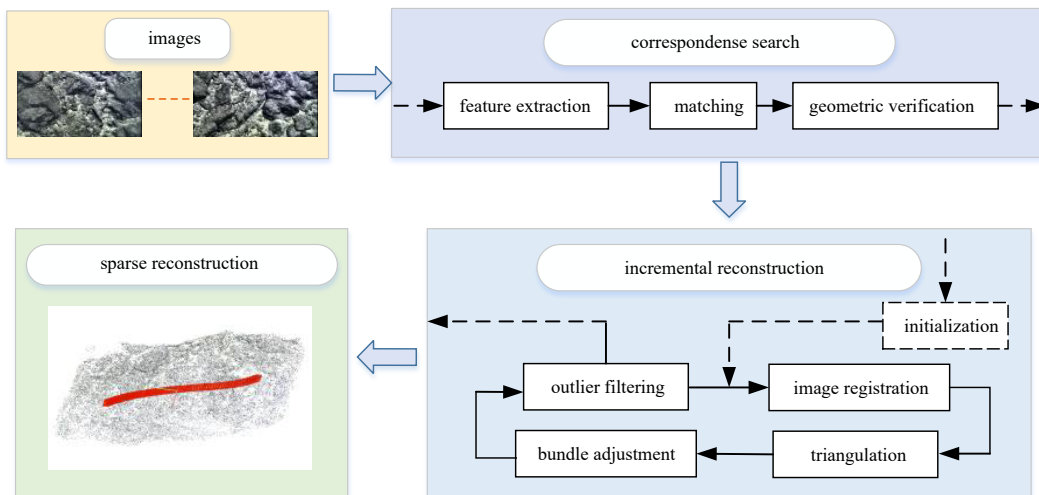


Fig. 4. Incremental SFM flow chart.

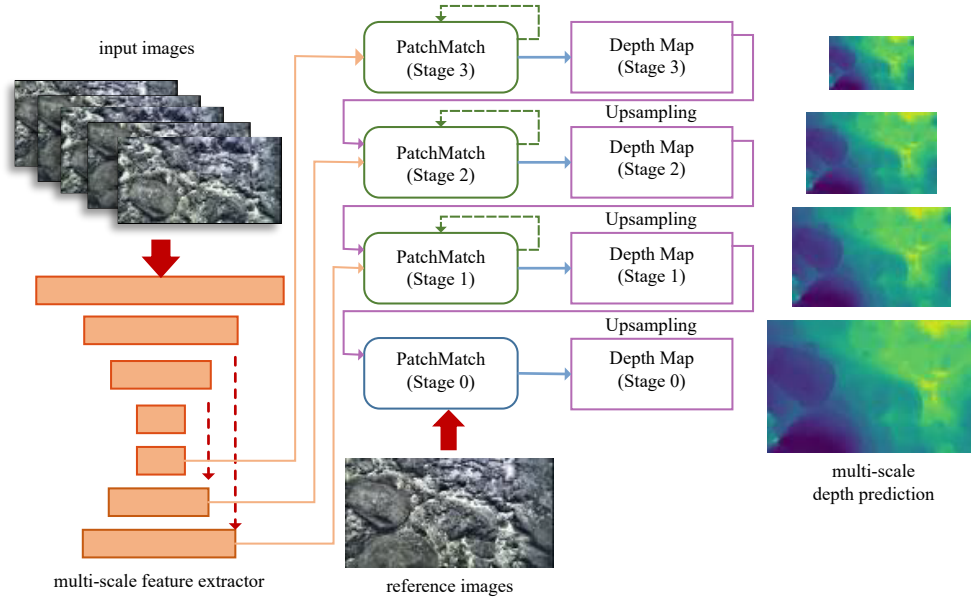


Fig. 5. Structure diagram of PatchmatchNet.

where $D_p(p)$ is the desired depth hypothesis, D is the depth map of the last iteration, x_i is the fixed offset, Δx_i is the additional offset to be learned, and N_p is the number of neighborhood pixels.

During the network training procedure, the backpropagation of parameters is accomplished by utilizing the loss function, which subsequently directs the optimization of model parameters. Equation (13) represents the loss function of PatchmatchNet, assuming n images are input:

$$\text{Loss}_{\text{total}} = \sum_{k=1}^3 \sum_{i=1}^{n_k} \text{Loss}_i^k + \text{Loss}_{\text{ref}}, \quad (13)$$

where the total loss function is divided into two parts. The first part is the difference Loss_i^k between the depth estimate of the i iteration of Patchmatch and the true value of the label in the k ($k = 1, 2, 3$) stages of the network. The second part is the difference Loss_{ref} between the optimized depth estimate and the true value. The total loss is obtained by adding the two parts.

After the depth map is obtained, fusing the camera pose and depth map information is necessary to convert the depth map to the point cloud. Assuming that the pixel coordinates of the image are (u, v) , D is the depth value of the pixel, the focal lengths of the camera are f_x and f_y , and the pixel coordinates of the optical center are u_0 and v_0 , Eq. (14) is used to calculate the coordinates X_c , Y_c , Z_c of the point cloud in the camera coordinate system:

$$\begin{cases} X_c = \frac{(u - u_0)D}{f_x}, \\ Y_c = \frac{(v - v_0)D}{f_y}, \\ Z_c = D. \end{cases} \quad (14)$$

Using the rotation matrix R and the translation vector

T of the camera external parameters, the point cloud coordinates X_w , Y_w , and Z_w in the world coordinate system are calculated by Eq. (15). The final 3D point cloud model is obtained after the point clouds generated by all depth maps are fused:

$$\begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}. \quad (15)$$

3 Results and discussion

3.1 Experimental environment and parameters

The hardware for this experiment consisted of an Intel(R) Core(TM) i5-12490 CPU processor, NVIDIA GeForce RTX3060 GPU, and 12 GB video memory. The other device is an Intel Core i5-12500H CPU model with 16 GB of RAM, NVIDIA GeForce RTX 3060 GPU, and 6 GB of video memory. The software is the Windows 10 operating system, Pytorch 1.13.0 deep learning framework, and PyCharm development platform.

The data for this paper was sourced from authentic video footage captured by the *Jiaolong* manned submersible on the sea floor, and the video parameters are shown in Table 1. The dataset was constructed following the process depicted in Fig. 2a under “Dataset Production”. Initially, key frames were extracted from the videos. Subsequently, the images were simply expanded through data augmentation techniques such as flipping and rotating. Finally, the images were annotated using the Labeling tool, yielding a total of 6 144 images along with their corresponding labels. The data was then partitioned into training, validation, and testing sets at a ratio of 7:1:2, respectively.

3.2 Algorithm performance analysis

3.2.1 Performance evaluation of image enhancement algorithms

To verify the effectiveness of the proposed image enhancement method, we compared the proposed algorithm with the multi-scale retinex with color restoration (MSRCR) algorithm (Jobson et al., 1997), the dark channel prior (DCP) algorithm (Yuan et al., 2020), and the adaptive gamma and contrast adjustment with weighted distribution (AGCWD) algorithm (Reinhard et al., 2001). The outcomes of the comparison are illustrated in Fig. 6.

Compared with alternative image enhancement algorithms, this approach substantially enhances both the contrast and color deviation of an image. The abundance of feature information in an image indicates its depth of texture and structure data, which is advantageous for image matching and recognition. The scale-invariant feature transform (SIFT) algorithm is a well-established method for extracting image features. The SIFT algorithm is utilized to derive image feature points prior to and subsequent to enhancement to assess the capability of the proposed enhancement method to recover image feature information. The experimental results are shown in Fig. 7.

After image enhancement using the proposed method, the number of feature points significantly increased compared with the original image, and the distribution in the area of interest of the image was denser.

To further verify the effectiveness of the proposed algorithm, we used the underwater image quality assessment metrics (UIQM) index of the human vision system

(Li et al., 2023) to evaluate the algorithm's performance. The comparative experimental results on the UIQM index are shown in Table 2. The proposed method performs well in enhancing deep-sea images. Specifically, regarding the UIQM index, compared with the MSRCR, DCP, and AGCWD algorithms, it increased by 73.65%, 103.81%, and 155.85%, respectively.

The above experimental results show that the proposed method can effectively enrich the feature information of the image, and is conducive to improving the accuracy of the subsequent target detection model and 3D reconstruction model.

3.2.2 Accuracy evaluation of megafaunal identification and counting algorithm

YOLOv5 comes in various sizes and complexities, including YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, and YOLOv5n. YOLOv5n, which has the smallest number of model parameters, was selected in this study for comparative experiments with similar algorithms to validate the effectiveness of the proposed method. The compared algorithms are Faster RCNN (Ren et al., 2017), YOLOv3 (Jiang et al., 2021), YOLOv7-Tiny (Ma et al., 2023), and YOLOv3-Tiny (Li and Liang, 2021). This study utilized the mean average precision (mAP) to measure the overall detection performance of the object detection algorithms, which is expressed as follows:

$$\text{mAP} = \frac{\sum_{n=1}^C \text{AP}(n)}{C} \times 100\%, \quad (16)$$

where AP refers to the average precision, and C denotes the number of categories. Two commonly used metrics are $\text{mAP}_{0.5}$ and $\text{mAP}_{0.5:0.95}$. $\text{mAP}_{0.5}$ is calculated by setting the IoU threshold to 0.5, meaning detection is consid-

Table 1. Video parameters

Resolution ratio	Frame rate format	Encoding method	Color space
1 920 × 1 080	25	H.264	RGB

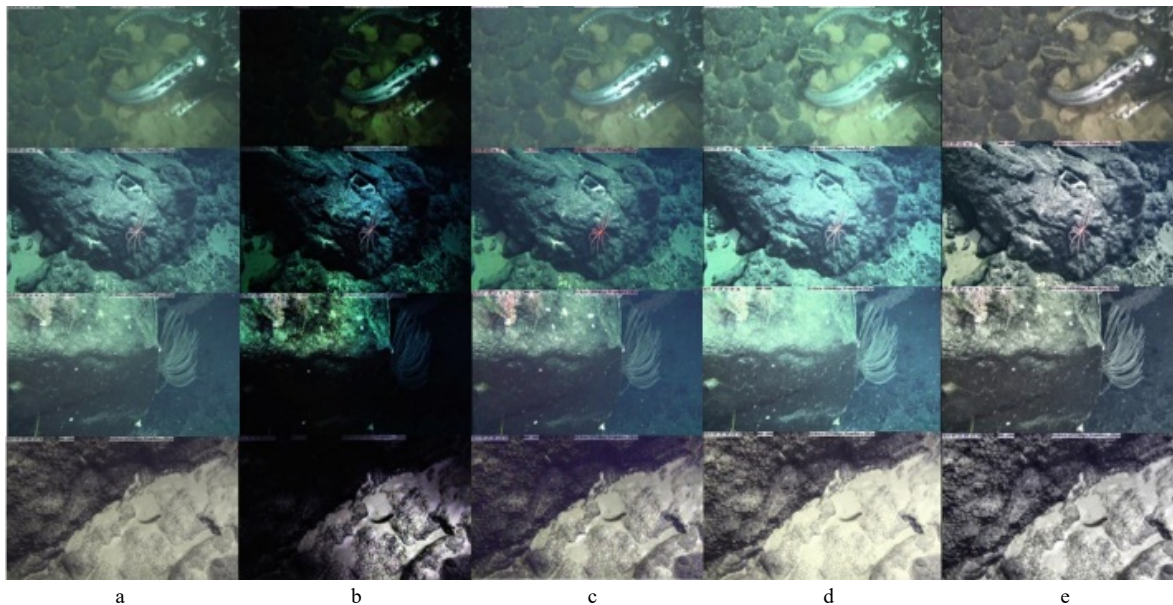


Fig. 6. Algorithm comparison results. a. Original drawing, b. MSRCR, c. DCP, d. AGCW, and e. this method.

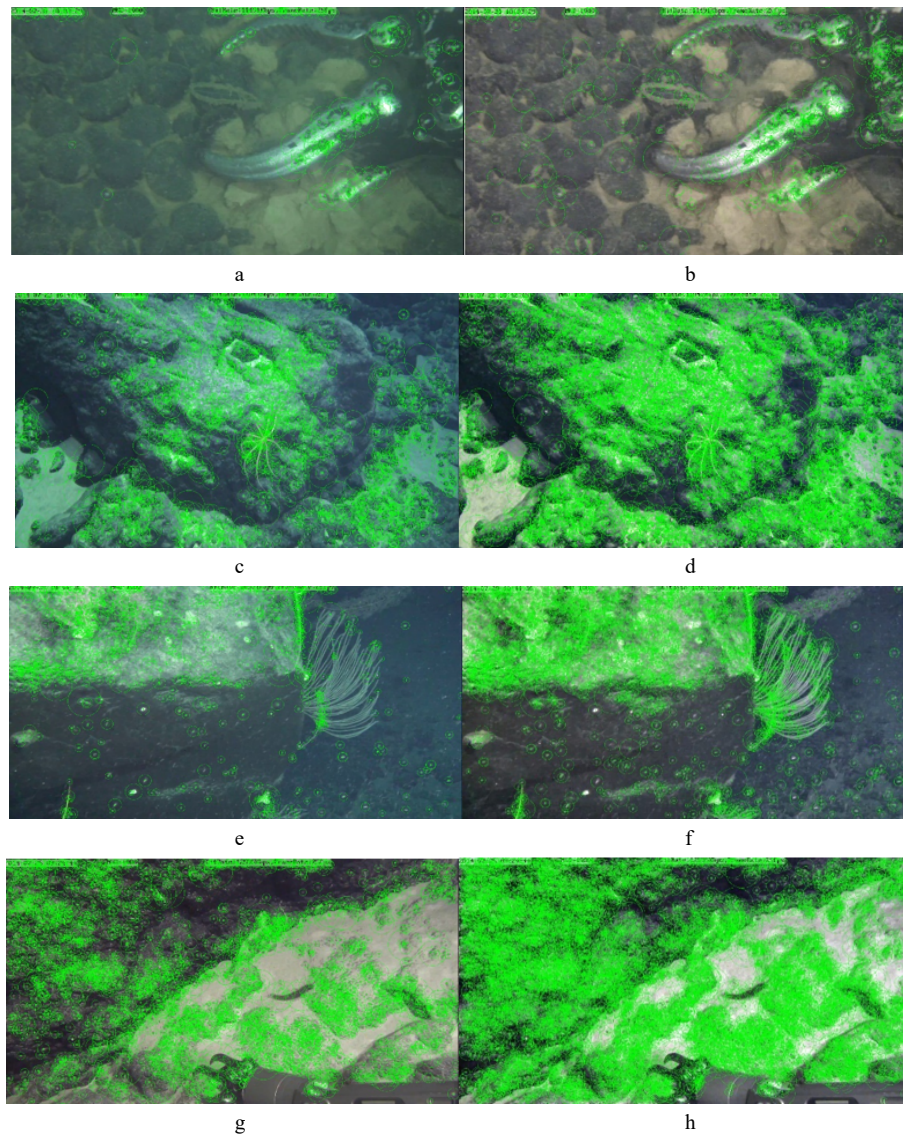


Fig. 7. SIFT feature detection effect visualization. a. Image 1, b. Image 1 with enhanced algorithm, c. Image 2, d. Image 2 with enhanced algorithm, e. Image 3, f. Image 3 with enhanced algorithm, g. Image 4 and h. Image 4 with enhanced algorithm (Table 2).

Table 2. Comparison of experimental results of the UIQM index

Image	Original	MSRCR	DCP	AGCW	This method
1	0.74	2.34	0.93	1.09	4.50
2	2.03	2.99	2.43	1.82	4.89
3	1.53	2.48	2.07	1.55	4.99
4	2.87	3.28	4.02	3.05	4.87
Average	1.80	2.77	2.36	1.88	4.81

ered a true positive when the IoU is above 0.5. $mAP_{0.5:0.95}$ is obtained by averaging the AP values over a range of IoU thresholds from 0.5 to 0.95 with increments of 0.5. The spatial complexity of an algorithm is measured by the number of parameters (Params).

To validate the effectiveness of the method proposed in this paper, the YOLOv5n model utilized herein was compared with Faster R-CNN, YOLOv3, YOLOv3-Tiny, and YOLOv7-Tiny algorithms on the dataset introduced

in this study. The specific results of the comparative experiments are presented in Table 3.

As indicated in Table 3, YOLOv5n has the fewest parameters among the compared algorithms, with only 1.8 million, yet it maintains a commendable detection accuracy. Its $mAP_{0.5}$ and $mAP_{0.5:0.95}$ are close to those of the most accurate YOLOv3. Therefore, considering the complexity of the model and its detection performance, the YOLOv5n model used in this paper offered superior accuracy with relatively low complexity. It was able to achieve high-precision detection of deep-sea organisms while maintaining fast inference speeds.

The recognition effect diagram of the YOLOv5 model is illustrated in Fig. 8. The figure illustrates the model's effective detection capabilities for megafauna. The taxa *Poliopogon* sp. and *Keratoisididae* sp. are enclosed in circles, with the target organism's species name and confidence level corresponding to each block. To further ana-

lyze the applicability of the YOLOv5 algorithm to the dataset and assess the accuracy of this algorithm in identifying the number of organisms, we detected and tested the number and category of organisms in 50 images using the YOLOv5 algorithm and a manual recognition method, respectively. The horizontal axis shows the serial number of the experimental images, and the vertical axis shows the number of creatures in the pictures. This paper presents an instance where the manual enumerating the same organisms' number and category coincides with the automated recognition of the said attributes.

The number contours of large deep-sea benthic organisms, as identified through the YOLOv5n algorithm and manual identification method, are depicted in Fig. 9. The

Table 3. Comparative experimental results of different algorithms

Model	mAP _{0.5} /%	mAP _{0.5:0.95} /%	Param/10 ⁶
Faster RCNN	67.9	44.4	137.1
YOLOv3	94.9	79.1	61.5
YOLOv3-Tiny	93.9	74.5	8.7
YOLOv7-Tiny	91.1	67.6	6.0
YOLOv5n	93.6	72.9	1.8

similarity in trend between the test value curve and ground truth curve suggests that the YOLOv5n algorithm accurately identifies many characteristics of benthic organisms in the deep sea. Specifically, the number of organisms counted manually is equivalent to the number of organisms identified automatically by the algorithm.

3.2.3 Performance evaluation of the 3D reconstruction algorithm

The objective of 3D reconstruction is to provide researchers with a visual representation of the topography of a specific region. To assess the viability and efficacy of this approach, we employed a subjective evaluation to appraise the reconstruction outcomes qualitatively.

In general, the production cost of datasets for multi-view 3D reconstruction methods based on deep learning is prohibitive. Therefore, we used the PatchmatchNet pre-trained model on the DTU ([Aanaes et al., 2016](#)) dataset to analyze the experimental data. The DTU dataset, published by the Technical University of Denmark, is an extensive resource for multi-view 3D reconstruction. Comprising 128 meticulously calibrated scenes, the dataset employs structured light scanning to capture objects from

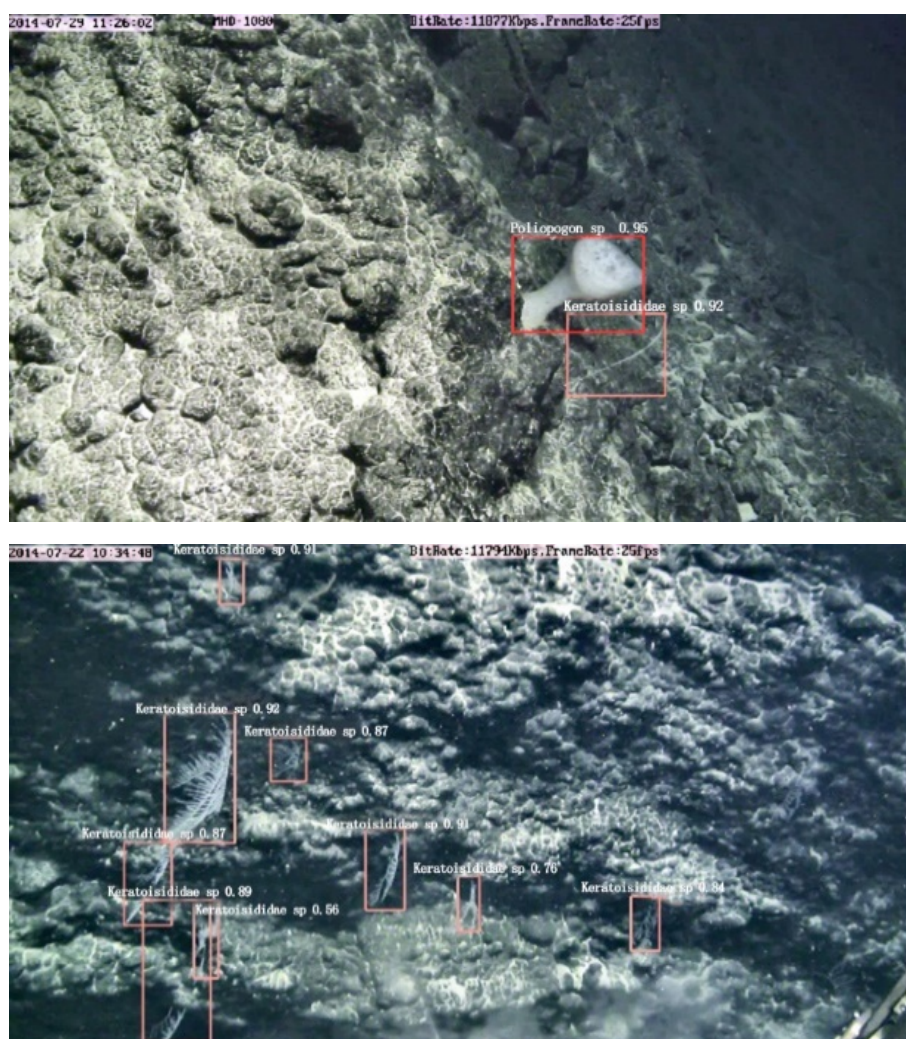


Fig. 8. The recognition effect of YOLOv5 at two locations in the Caiwei Guyot area.

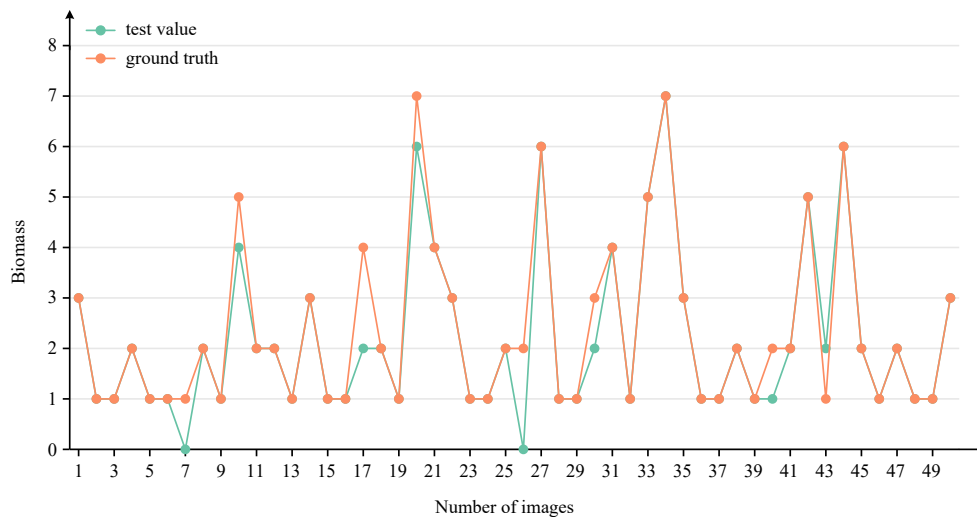


Fig. 9. Comparison of YOLOv5n test value and ground truth.

multiple perspectives as a benchmark within the domain of multi-view 3D reconstruction.

Figure 10 presents the ground truth point cloud models of four distinct scenes from the DTU dataset, the point cloud models generated by COLMAP (Schönberger and Frahm, 2016), and those produced by PatchmatchNet. COLMAP, an open-source tool for 3D reconstruction, serves as a benchmark in the field of multi-view 3D reconstruction. As depicted in Fig. 10, the point cloud models reconstructed by PatchmatchNet exhibit greater integrity and more detailed recovery compared to those of COLMAP, closely approximating the ground truth models. This comparison underscores the effectiveness of PatchmatchNet.

The dense point cloud model produced by this method after processing experimental data is shown in Fig. 11. The reconstructed results are incredibly comprehensive, restoring the majority of the topographic and geomorphic data for the region. Figure 11a indicates that, despite the challenges posed by minor oscillations caused by water flow and the uniformity of surface color for feature extraction, our algorithm can still effectively reconstruct the sponge, and the resulting model possesses a high degree of recognizability. By rotating the 3D model, researchers

can conduct a more comprehensive analysis of the cliff's steepness. The alternating distribution of fine gravel and gravel blocks in the area is illustrated in Fig. 11b. The 3D model is utilized to depict the extent of gravel protrusion,

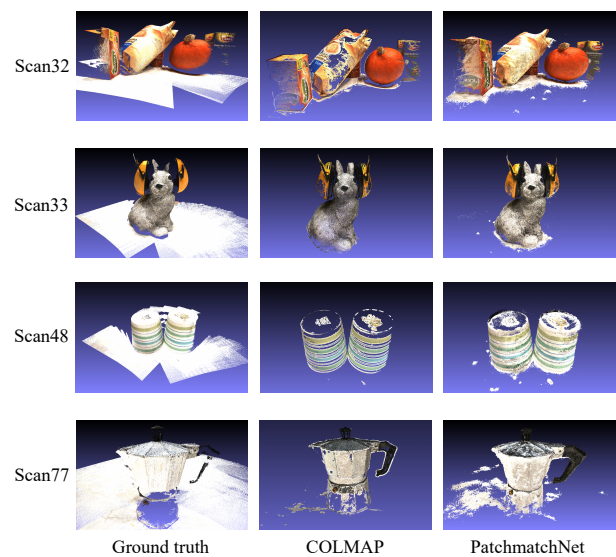


Fig. 10. Comparison between the ground truth and the reconstructed point cloud.

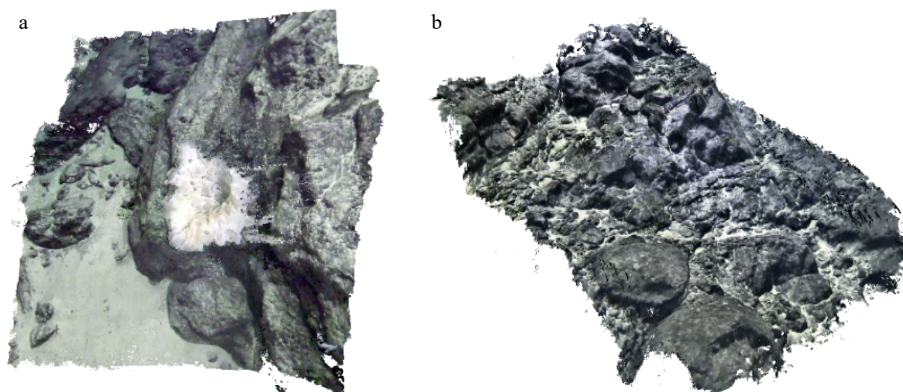


Fig. 11. Part of the microgeomorphic 3D reconstruction result.

enabling researchers to comprehend the microgeomorphic information more intuitively. The experimental results validate the viability of the reconstruction method and show the robustness of PatchmatchNet across various application scenarios.

3.3 Species composition

The megafaunal identification and counting algorithm based on YOLOv5 identified ten classes of organisms and six phyla in the Caiwei Guyot region. The specific findings of the survey are presented in Table 4. As shown in the table, the proportion of Hexactinellida that comprised the majority of this survey was 32.56%. Following this, the Anthozoa comprised 16.73%. The least common animal in this survey is Polychaeta. Furthermore, in adherence to the class classification standard, we constructed the bar chart shown in Fig. 12 to provide a more direct visualization of the organism composition.

3.4 Main groups and dominance

In this study, McNaughton (Chen et al., 2009) (Y) was used to reflect the dominance of an organism. Y represents the degree of dominance of a particular species in the community, and species with $Y > 0.02$ are dominant species in the community. The statistical results are shown in Table 5. Based on the results in the table, the most dominant group of Hexactinellida is followed by Anthozoa, and the most vulnerable group is Echinoidea. Similarly, we normalized the dominance and produced a normalized dominance pie chart, as depicted in Fig. 13, from which the same conclusion can be observed.

3.5 Spatial distribution characteristics

The biological distribution was analyzed based on various regions, including the south peak, the south slope, the western foothills, the north peak, the north mountain slope, the west side of the seamount, and the northern foothills. A histogram depicting the distribution of organisms in these areas was generated as a result of the analysis, as shown in Fig. 14.

South peak (75th dive): The south peak includes Asteroidea, Chordata, Malacostraca, Anthozoa, and others. The organisms that were observed are dispersed, lacking any conspicuous high-value regions.

South slope (76th dive): The south slope revealed several significant groups, including Porifera and Chordata, followed by Asteroidea, Crinoidea, Holothuroidea, and others. Furthermore, microscopic Malacostraca and Anthozoa were observed, and the species diversity was greater than that observed on the opposite side of the mountain. Numerous organisms are present, with Porifera and Actinopterygii constituting the majority. Chordata and Malacostraca are prevalent, whereas Porifera, Holothuroidea, Asteroidea, and Crinoidea are uncommon.

Western foothills (77th dive): There is less biodiversity and diversity compared with the tops and slopes of the

Table 4. Species composition

Phylum	Class	Order/family/genus	Number		
Porifera	Demospongiae	Callyspongiidae sp.	41		
		Cladorhizidae sp.	3		
		<i>Chondrocladia</i> sp.	3		
	Hexactinellida	<i>Caulophacus</i> sp.	16		
		<i>Poliopogon</i> sp.	19		
		<i>Hyalonema</i> sp.	12		
		Corbitellinae sp.	17		
		<i>Bolosoma</i> sp.	5		
		<i>Rhizophyta</i> sp.	4		
		<i>Tretopleura</i> sp.	3		
		<i>Saccocalyx</i> sp.	1		
		<i>Semperella</i> sp.	106		
Cnidaria	Anthozoa	<i>Chrysogorgia</i> sp.	5		
		<i>Fungiacyathus stephanus</i>	1		
		Primnoidae sp.	14		
		<i>Actinoscyphia</i> sp.	1		
		Keratoisididae sp.	56		
		<i>Iridogorgia</i> sp.	6		
		<i>Bathypathes</i> sp.	3		
		Actinostolidae sp.	3		
		Actiniaria sp.	5		
		Annelida	Polychaeta	Phyllodocida sp.	1
				Polynoidae sp.	2
		Arthropoda	Malacostraca	<i>Heterocarpus</i> sp.	17
Aristeidae sp.	18				
<i>Nematocarcinus</i> sp.	4				
Echinodermata	Crinoidea	Crinoidea sp.	9		
		Hyocrinidae sp.	3		
		Bathycrinidae sp.	4		
		Pentametrocrinidae sp.	7		
		Asteroidea	<i>Freyastera</i> sp.	7	
	Pterasteridae sp.		1		
	Asteroidea sp.		1		
	Freyellidae sp.		8		
	<i>Freyastera basketa</i>		5		
	Brisingidae sp.		2		
	<i>Freyastera mortenseni</i>		1		
	Echinoidea		Echinoidea sp.	8	
			Holothuroidea	<i>Paroriza</i> sp.	1
	<i>Hansenothuria</i> sp.	2			
	<i>Paelopatides</i> sp.	32			
	Synallactidae sp.	13			
	<i>Psychropotes</i> sp.	11			
	Elpidiidae sp.	5			
	<i>Enypniastes</i> sp.	2			
	<i>Molpadidemas</i> sp.	4			
	Holothuroidea sp.	1			
	<i>Peniagone</i> sp.	1			
	<i>Benthodytes</i> sp.	3			
	Chordata	Actinopterygii	<i>Aldrovandia afinis</i>	10	
			Ophidiiformes sp.	9	
			Synaphobranchidae sp.	13	
Halosauridae sp.			9		
Ophidiidae sp.			15		
Macrouridae sp.			6		
<i>Abyssoberyx</i> sp.			3		

seamount. The main groups are Porifera and Holothuroidea, Chordata, Malacostraca, Anthozoa, and Asteroidea. These are small individual organisms with low abundance and sporadic distribution.

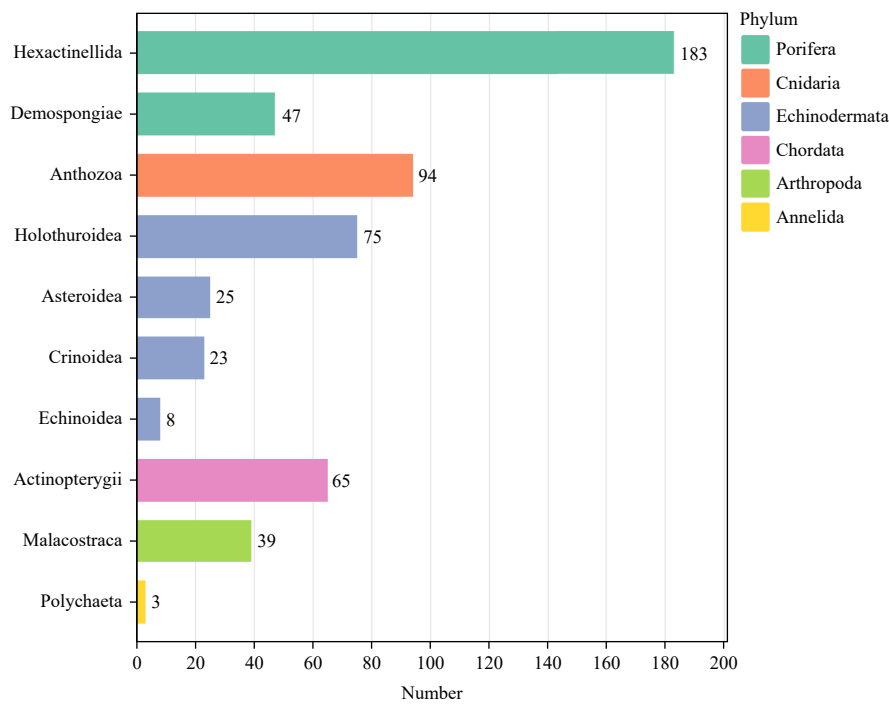


Fig. 12. Number of main groups.

North peak (78th dive): The most important groups are Anthozoa, Holothuroidea, and Chordata, which are diverse and relatively abundant. In addition, Porifera, Crinoidea, and Malacostraca classes are common. At 2 400–2 250 m, mainly Porifera and Anthozoa are present. At 2 250–1 900 m, Chordata and Holothuroidea are the main organisms. From 1 900 m to 1 850 m, there are few organisms, mainly Anthozoa, Porifera, and Asteroidea.

The first section of the north slope (79th dive): The main groups are Anthozoa and Porifera, in addition to Asteroidea, Crinoidea, Malacostraca, Holothuroidea, Chordata, and other groups. Anthozoa is the most abundant and occurs in groups on overhanging banks and boulders.

The second section of the north slope (80th dive): Porifera is the main group in this region. In addition, there are Anthozoa, Asteroidea, Holothuroidea, Crinoidea,

Table 5. Main groups and dominance

Class	Degree of dominance
Demospongiae	0.005
Hexactinellida	0.05
Anthozoa	0.03
Polychaeta	0.004 3
Malacostraca	0.008 2
Crinoidea	0.006 4
Asteroidea	0.007
Echinoidea	0.002
Holothuroidea	0.008
Actinopterygii	0.01

Malacostraca, and Chordata. Due to the depth of the water, the number of giant benthos is small, sporadic, and occasionally observed attached to some Porifera organisms, such as anemones and Crinoidea.

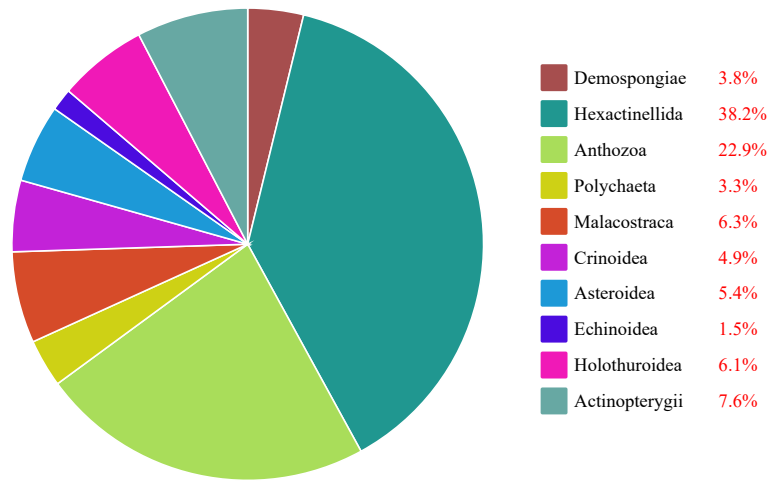


Fig. 13. The percentage distribution of main groups.

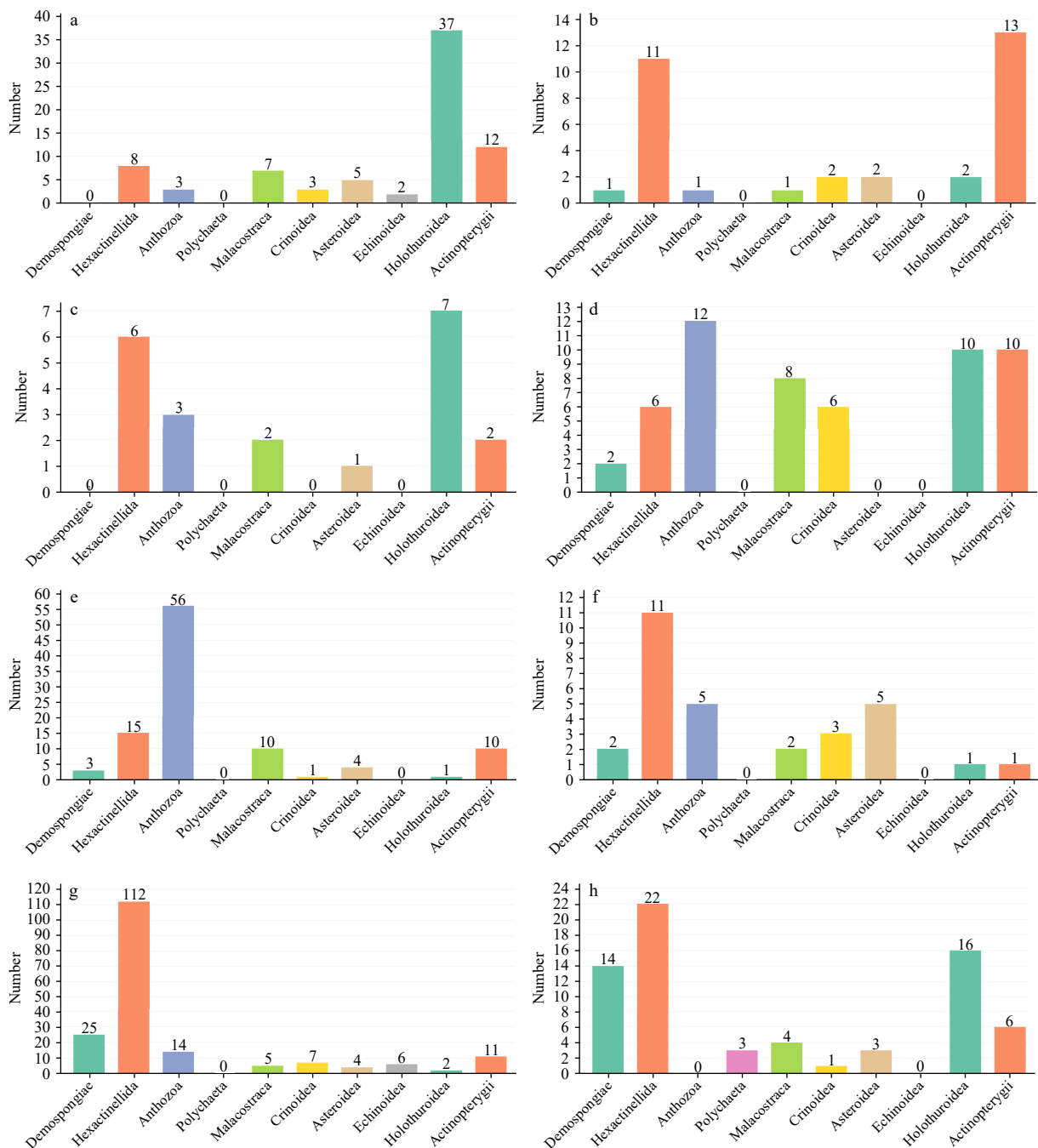


Fig. 14. Number of main groups in each region. a. South peak, b. south slope, c. western foothills, d. north peak, e. first section of the north slope, f. second section of the north slope, g. west side of the seamount, and h. northern foothills.

West side of the seamount (81st dive): Porifera is the main group, with many organisms observed. In addition to Porifera, Anthozoa, Asteroidea, Crinoidea, Holothuroidea, Echinoidea, Malacostraca, and Chordata are present. From 2 100 m to 1 880 m, the sediment is primarily rocky, covered with more sediments, and the number of organisms is small, mainly Porifera. From 1 880 m to 1 570 m, there were sheets of crust, and many organisms, especially Porifera, appeared in groups at the cliff. There were more than 20 Porifera in one field of view. In addition to Porifera, we observed Anthozoada, Asteroidea, Crinoidea, and Chordata.

Northern foothills (82nd dive): There are more species of Porifera and Holothuroidea, and the biodiversity is significantly greater than in the western foothills. In addition, there are Asteroidea, Chordata, Red Malacostraca, armor Malacostraca, and others. Porifera and Holothuroidea are more abundant than in the western foothills.

3.6 Geomorphologic features

Assisted by the designed 3D reconstruction algorithm, it was found that the Caiwei Guyot is roughly elliptical in shape, with a long axis from northeast to southwest. It is approximately 130 km long and consists of broad and

gentle summit platforms and steep slopes, with a total area of approximately 11 516 km². The depth of the foot of the seamount is 5 500 m, and it develops into two flat-top seamounts at a shallow depth of 3 000 m. Among them, the Caiwei Guyot is a large flat-top mountain in the north-east, with a water depth of 1 500–1 650 m at the edge of the summit platform. The Caiwei Guyot is a small flat-top mountain in the southwest, with a water depth of 1 850–1 900 m at the edge of the summit platform. The secondary landform of the seamount includes three types: a summit platform, hillside, and intermountain valley. The two hilltop platforms are very flat, with an area of approximately 1 912 km² and 125 km². The overall topography of the sea slope is steep, and its three-level landform can be divided into four types: gentle slope, slope, steep slope, and precipitous slope. Many small ridges and hills developed.

By leveraging 3D reconstruction technology, a more intuitive analysis of the seamount's geomorphologic characteristics can be achieved. For instance, Fig. 15 depicts a 3D model of a region situated on the southern slope of the Caiwei Guyot, where the water profundity ranges from approximately 1 530 m to 1 556 m. The model predicts the presence of a mixture of sizable and minute gravel inclusions within this region, with a reduced quantity of sediment. The water depths for Figs 15a, b, and c are approximately 1 556 m to 1 530 m, 1 566 m to 1 556 m, and 1 825 m to 1 566 m, respectively. The landforms of this area are comparatively level, consisting primarily of plate and gravel crusts with minimal sediment cover, as illustrated in Fig. 15a and Fig. 15b. With the transition from plate crust to gravel crust, the topography in Fig. 15c is

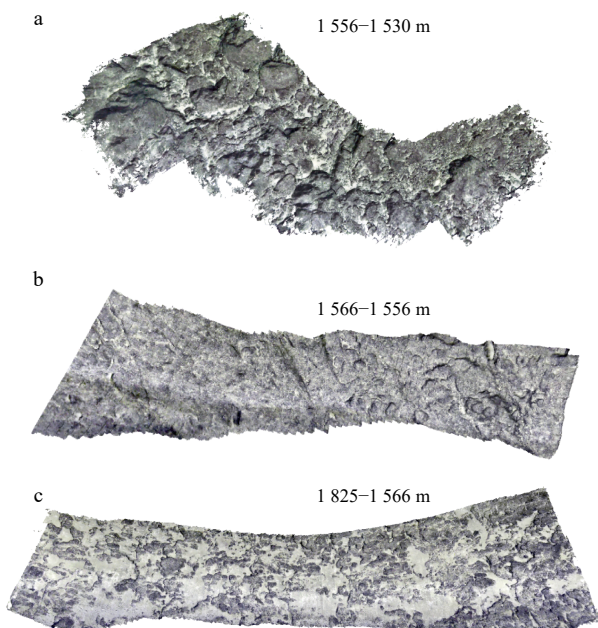


Fig. 15. 3D reconstruction results of part of the Caiwei Guyot area. The water depths in a, b, and c are approximately 1 556 m to 1 530 m, 1 566 m to 1 556 m, and 1 825 m to 1 566 m, respectively.

relatively flat, and the distribution range is relatively restricted. There is an increase in sediment.

Generally, the summit exhibits a level topography, entirely enveloped in sediment and devoid of any exposed crust. The plate is a predominant crust type on the steepest slope, where the crust is more continuous, and a portion of the crust surface is covered with sediment. A marginal quantity of plate crust may intermittently manifest along the slope characterized by a slight gradient, where sediment is prevalent. Cobalt-rich nodules appear on the steep hillside between 3 500 m and 4 000 m. Sediment covers the outcrop, rendering its appearance and form obscure. At times, the crust manifests itself, and the substructure becomes apparent.

4 Conclusions

This study investigated the megafauna and landform characteristics of the Caiwei Guyot region via data-driven technology. In summary, the findings of this research are as follows.

(1) We developed a megafaunal identification and counting algorithm based on YOLOv5, specifically tailored to identify and enumerate megafauna in the Caiwei Guyot region. This algorithm effectively categorizes and tallies megafauna with precision, significantly enhancing the efficiency of analyses related to biological species composition, dominance, and distribution patterns.

(2) Our research has yielded a 3D reconstruction algorithm for landforms utilizing PatchmatchNet. This algorithm accurately captures the terrain's characteristics in a 3D format, thus improving researchers' ability to intuitively grasp and comprehensively analyze the area's topographical features.

(3) From a novel perspective, our study accurately and efficiently unveiled the unique geomorphic features and the rich biodiversity of megafauna in the Caiwei Guyot area, contributing to intelligent deep-sea biological and geological research advancement.

(4) The Caiwei Guyot area exhibits an exceptional level of biodiversity. Hexactinellida is the most dominant group of giant benthos on the seamount, characterized by many abundant populations, accounting for 32.56% of the total number of organisms counted in this study. They predominantly inhabit the seamount's west, south, and north sides. Anthozoa constitute the second most dominant group, accounting for 16.73% of the total number of organisms counted in this study, primarily found on the eastern and northern slopes. Other significant groups include Holothuroidea, Actinopterygii, and Malacostraca.

(5) The overall steep sea slope is punctuated by numerous small ridges and hills, with the terrain further classified into gentle, steep, and precipitous slopes. The summit platform of the Guyot features an extremely flat topography, contrasting with the surrounding slopes.

References

- Aanæs H, Jensen R R, Vogiatzis G, et al. 2016. Large-scale data for multiple-view stereopsis. *International Journal of Computer Vision*, 120(2): 153–168, doi: [10.1007/s11263-016-0902-9](https://doi.org/10.1007/s11263-016-0902-9)
- Ancuti C O, Ancuti C, De Vleeschouwer C, et al. 2018. Color balance and fusion for underwater image enhancement. *IEEE Transactions on Image Processing*, 27(1): 379–393, doi: [10.1109/TIP.2017.2759252](https://doi.org/10.1109/TIP.2017.2759252)
- Baker M R, Williams K, Greene H G, et al. 2021. Use of manned submersible and autonomous stereo-camera array to assess forage fish and associated subtidal habitat. *Fisheries Research*, 243: 106067, doi: [10.1016/j.fishres.2021.106067](https://doi.org/10.1016/j.fishres.2021.106067)
- Barnes C, Shechtman E, Finkelstein A, et al. 2009. PatchMatch: a randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (TOG)*, 28(3): 24
- Catmull E E. 1974. *A Subdivision Algorithm for Computer Display of Curved Surfaces*. Springfield: NTIS
- Chen Jiazhang, Meng Shunlong, You Yang, et al. 2009. Characteristic of phytoplankton community in Lake Wuli, Lake Taihu. *Ecology and Environmental Sciences (in Chinese)*, 18(4): 1358–1367, doi: [10.16258/j.cnki.1674-5906\(2009\)04-1358-10](https://doi.org/10.16258/j.cnki.1674-5906(2009)04-1358-10)
- Collot J Y, Lallemand S, Pelletier B, et al. 1992. Geology of the d'Entrecasteaux-New Hebrides arc collision zone: results from a deep submersible survey. *Tectonophysics*, 212(3–4): 213–217, 221–241
- Dai Jifeng, Qi Haozhi, Xiong Yuwen, et al. 2017. Deformable convolutional networks. In: *Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 764–773
- Ding Zhongjun. 2021. *Operational Techniques for Deep-Sea Manned Submersible Exploration (in Chinese)*. Harbin: Harbin Engineering University Press, 26–35
- Ding Zhongjun, Liu Chen, Li Dewei, et al. 2023. Deep-sea biological detection method based on lightweight YOLOv5n. *Sensors*, 23(20): 8600, doi: [10.3390/s23208600](https://doi.org/10.3390/s23208600)
- Duan Xiaoxu, Duan Jun, Zhang Xiangxiang. 2023. Study on the typical micro-geomorphologic characteristics of seabed in a certain sea area in southern Zhuhai. *Guizhou Science (in Chinese)*, 41(3): 59–62, 91
- Geyer R A. 1977. Submersibles and their use in oceanography and ocean engineering. *Ocean Management*, 3(2): 137–148, doi: [10.1016/0302-184X\(77\)90012-9](https://doi.org/10.1016/0302-184X(77)90012-9)
- Grinyó J, Francescangeli M, Santin A, et al. 2022. Megafaunal assemblages in deep-sea ecosystems of the Gulf of Cadiz, northeast Atlantic Ocean. *Deep-Sea Research Part I: Oceanographic Research Papers*, 183: 103738, doi: [10.1016/j.dsr.2022.103738](https://doi.org/10.1016/j.dsr.2022.103738)
- Guo Binbin, Wang Weiqiang, Shu Yeqiang, et al. 2020. Observed deep anticyclonic cap over Caiwei Guyot. *Journal of Geophysical Research: Oceans*, 125(10): e2020JC016254, doi: [10.1029/2020JC016254](https://doi.org/10.1029/2020JC016254)
- Hou Zhongsheng, Xu Jianxin. 2009. On data-driven control theory: the state of the art and perspective. *Acta Automatica Sinica (in Chinese)*, 35(6): 650–667, doi: [10.3724/SP.J.1004.2009.00650](https://doi.org/10.3724/SP.J.1004.2009.00650)
- Jiang Xinbei, Gao Tianhan, Zhu Zichen, et al. 2021. Real-time face mask detection method based on YOLOv3. *Electronics*, 10(7): 837, doi: [10.3390/electronics10070837](https://doi.org/10.3390/electronics10070837)
- Jobson D J, Rahman Z U, Woodell G A. 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7): 965–976, doi: [10.1109/83.597272](https://doi.org/10.1109/83.597272)
- Jocher G. 2020. Ultralytics YOLOv5. <https://docs.ultralytics.com/zh/models/yolov5/#citations-and-acknowledgements> [2020-06-26/2023-03-16]
- Li Lifu, Liang Yi. 2021. Deep learning target vehicle detection method based on YOLOv3-tiny. In: *Proceedings of 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference*. Chongqing: IEEE, 1575–1579
- Li Dong, Zhao Jun, Liu Chenggang et al. 2018. Advances of living environment characteristics and biogeochemical processes in the hadal zone. *Earth Science (in Chinese)*, 43(S2): 162–178
- Li Ting, Zhou Xianchun, Zhang Ying, et al. 2023. Underwater image enhancement based on IMSRCR and CLAHE-WGIF. *Instrumentation*, 10(2): 19–29
- Liu Rulong, Wang Li, Wei Yuli, et al. 2018. The hadal biosphere: recent insights and new directions. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 155: 11–18, doi: [10.1016/j.dsr2.2017.04.015](https://doi.org/10.1016/j.dsr2.2017.04.015)
- Liu Zhehao, Zhang Jianxing, Wang Bing, et al. 2023. Geomorphological characteristics and geological processes of Caroline M4 guyot. *Oceanologia et Limnologia Sinica (in Chinese)*, 54(2): 351–361
- Ma Li, Zhao Liya, Wang Zixuan, et al. 2023. Detection and counting of small target apples under complicated environments by using improved YOLOv7-tiny. *Agronomy*, 13(5): 1419, doi: [10.3390/agronomy13051419](https://doi.org/10.3390/agronomy13051419)
- Ottesen D, Dowdeswell J A, Bellec V K, et al. 2017. The geomorphic imprint of glacier surges into open-marine waters: examples from eastern Svalbard. *Marine Geology*, 392: 1–29, doi: [10.1016/j.margeo.2017.08.007](https://doi.org/10.1016/j.margeo.2017.08.007)
- Perez J A A, Kitazato H, Sumida P Y G, et al. 2018. Benthopelagic megafauna assemblages of the Rio Grande rise (SW Atlantic). *Deep-Sea Research Part I: Oceanographic Research Papers*, 134: 1–11, doi: [10.1016/j.dsr.2018.03.001](https://doi.org/10.1016/j.dsr.2018.03.001)
- Pizer S M, Amburn E P, Austin J D, et al. 1987. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3): 355–368
- Reinhard E, Adhikhmin M, Gooch B, et al. 2001. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5): 34–41
- Ren Shaoqing, He Kaiming, Girshick R, et al. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis*

- and Machine Intelligence, 39(6): 1137–1149, doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)
- Schönberger J L, Frahm J M. 2016. Structure-from-motion revisited. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 4104–4113
- Usui A, Nishimura A, Iizasa K. 1993. Submersible observations of manganese nodule and crust deposits on the tenpo seamount, northwestern Pacific. *Marine Georesources & Geotechnology*, 11(4): 263–291
- Usui A, Sato H, Nishi K, et al. 2013. Geological characterization of co-rich ferromanganese crusts over the northwestern Pacific seamounts. In: Proceedings of 2013 OCEANS. San Diego: IEEE, 1–3, doi: [10.23919/OCEANS.2013.6741019](https://doi.org/10.23919/OCEANS.2013.6741019)
- Wang Fangjinhua, Galliani S, Vogel C, et al. 2021. Patchmatch-Net: learned multi-view patchmatch stereo. In: Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 14189–14198
- Wang Chunzhi, Niu Hongxia. 2022. Sand-dust degraded image enhancement algorithm based on histogram equalization and MSRCR. *Computer Engineering* (in Chinese), 48(9): 223–229, doi: [10.19678/j.issn.1000-3428.0062764](https://doi.org/10.19678/j.issn.1000-3428.0062764)
- Xie Jun, Di Jianglei, Qin Yuwen. 2022. Application of deep learning in underwater imaging (invited). *Acta Photonica Sinica* (in Chinese), 51(11): 1101001, doi: [10.3788/gzxb20225111.1101001](https://doi.org/10.3788/gzxb20225111.1101001)
- Yuan Yubin, Shen Yu, Peng Jing, et al. 2020. Defogging technology based on dual-channel sensor information fusion of near-infrared and visible light. *Journal of Sensors*, 2020(1): 8818650
- Zhang Jianxing, Song Yongdong, Luan Zhendong, et al. 2021. Analysis of the characteristics of submarine topography and distribution of sediments near Juehua Island, Liaodong Bay. *Marine Sciences* (in Chinese), 45(9): 40–47