

Article

Morphometric Information for Yangtze Finless Porpoises Using Detection-Guided SAM2 Segmentation with UAV Imagery

Dongxu Yang^{1,2}, Xirui Xu³, Shengmao Zhang¹, Zuli Wu¹ , Tianfei Cheng¹ , Jianglong Que⁴, Siyao Wu¹  and Fei Wang^{1,*} 

¹ Key Laboratory of Fisheries Remote Sensing, Ministry of Agriculture and Rural Affairs, East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200090, China; m240701892@st.shou.edu.cn (D.Y.); zhangsm@ecsf.ac.cn (S.Z.); wuzl@ecsf.ac.cn (Z.W.); chengtf@ecsf.ac.cn (T.C.)

² College of Information Technology, Shanghai Ocean University, Shanghai 201306, China

³ College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China; 242142094@st.usst.edu.cn

⁴ Aquatic Conservation and Rescue Center of Jiangxi Province, Nanchang 330096, China

* Correspondence: wangfei@ecsf.ac.cn

Abstract

Morphometric information provides quantitative descriptors of cetacean size and shape and may support future assessments of individual condition and population status when combined with appropriate biological calibration. However, conventional contact-based measurements are difficult to apply to free-ranging Yangtze finless porpoises (*Neophocaena asiaeorientalis*). UAV imagery offers a non-contact means of acquiring porpoise morphometric data, but automated workflows for converting UAV observations into reliable body-surface measurements remain limited. To address this gap, an oriented detection-guided SAM2 morphometric workflow, termed ODG-SAM2-Morph, was developed to automatically extract body-surface morphometric parameters from UAV imagery. The workflow integrates YOLO26-OBb for oriented target localization, SAM2 for prompt-guided body-surface segmentation, and differentiated morphometric extraction strategies for complete-body and partial-body samples. Guided by oriented detections, the small SAM2 model with the R-Box + 1FG prompt generated body-surface masks with mean IoU, mean Dice, Precision, and Recall values of 0.824, 0.901, 0.883, and 0.932, respectively. For complete-body samples, automatically extracted body-length and body-width parameters showed preliminary agreement with manual measurements. For 114 partial-body samples, the automatically extracted maximum visible body width agreed reasonably well with manual measurements, with a MAE of 3.32 cm and a MAPE of 11.93%. Continuous-sequence analysis further showed that the representativeness of maximum visible body width depended on trunk exposure, contour clarity, body posture, and inter-frame stability. By converting UAV observations into quantitative image-derived morphometric measurements, ODG-SAM2-Morph provides a practical basis for non-contact morphometric monitoring of Yangtze finless porpoises under natural survey conditions.



Academic Editor: Yang Liu

Received: 5 August 2026

Revised: 30 August 2026

Accepted: 8 September 2026

Published: 9 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: Yangtze finless porpoise; UAV imagery; deep learning; oriented object detection; SAM2; morphometric measurement; ecological monitoring

Key Contribution: A UAV-based framework integrating YOLO26-OBb and SAM2 was developed for non-contact morphometric measurement of Yangtze finless porpoises. The framework enables oriented target localization, pixel-level body-surface segmentation, and morphometric extraction from both complete-body and partial-body UAV images.

1. Introduction

Morphometric information provides quantitative descriptors of individual size and shape and may contribute to the long-term monitoring and conservation of the Yangtze finless porpoise (*Neophocaena asiaeorientalis*). Body length, maximum body width, and longitudinal width profiles are commonly used to describe cetacean body size and shape. When calibrated against appropriate biological reference data, these parameters may also provide inputs for assessing growth and body condition as well as for supporting long-term population monitoring [1,2]. The Yangtze finless porpoise is an endemic freshwater cetacean inhabiting the Yangtze River basin [3,4]. The Yangtze finless porpoise is currently listed as a Class I nationally protected wild animal in China and as Critically Endangered on the IUCN Red List [5,6]. As a top predator in the Yangtze River aquatic ecosystem, the Yangtze finless porpoise is closely associated with prey availability and is responsive to hydrological variation and anthropogenic disturbance, making it a valuable indicator of aquatic habitat health [7]. The Yangtze finless porpoise's population status therefore serves as an important indicator of the ecological health of the Yangtze River aquatic ecosystem. However, conventional morphometric measurements in finless porpoises have mainly relied on contact-based methods or close-range observations, often involving capture, restraint, or intensive human disturbance [8]. Contact-based approaches are costly, potentially stressful to animals, and difficult to apply repeatedly over long periods and large spatial scales [9,10]. Therefore, low-disturbance, repeatable, and automated approaches are needed to support long-term morphometric monitoring of free-ranging Yangtze finless porpoises.

Unmanned aerial vehicle (UAV) platforms provide a new approach for the non-contact acquisition of body morphological parameters in the Yangtze finless porpoise [11,12]. Compared with close-range boat-based observations and manual contact-based measurements, UAVs offer several advantages, including low disturbance, rapid deployment, wide spatial coverage, and high-resolution imaging [13]. In particular, UAV-based nadir-view imaging with relatively stable geometry can directly record the body-axis orientation and water-surface visible contour of porpoises during natural surfacing events, providing an image basis for extracting individual-scale morphological information [14,15]. The central challenge is to convert UAV-acquired remote-sensing data into ecologically useful information. The Yangtze finless porpoise is rare and endangered, and its population density in natural waters is relatively low [16]. Field UAV surveys therefore usually require large-area image acquisition to increase the probability of recording porpoise individuals. Many images do not contain porpoise targets. Porpoises, when present, often occupy only a small proportion of the image and occur within complex water-surface backgrounds, including ripples, sun glint, turbid water, and splashes [17]. In addition, porpoises are exposed at or near the water surface only briefly during surfacing, rolling, or swimming, which requires high-frequency image acquisition to capture measurable body postures and visible contours [18,19]. Large-area, high-frequency image acquisition substantially increases image volume and the difficulty of target screening and sample analysis [20]. Therefore, automatic target detection and robust visible-contour extraction are essential for converting UAV imagery into individual-scale morphometric information.

Deep learning-based object detectors, especially the YOLO series, provide efficient tools for automatic animal target recognition in UAV ecological imagery because of their high detection speed and adaptability to complex backgrounds [21,22]. Such detectors can rapidly screen large-format UAV images, reduce manual inspection, and provide target position, scale, and orientation information for subsequent fine-scale analysis. Conventional task-specific segmentation models, however, generally require extensive high-quality pixel-level annotations [23,24]. Segment Anything Model 2 (SAM2) offers strong zero-shot segmentation capability and can be applied to new target categories without task-specific

training, making it a promising tool for extracting porpoise body-surface masks when annotated UAV datasets are limited [25,26].

UAV-based cetacean studies have already demonstrated the feasibility of target detection, species identification, photogrammetric measurement, and increasingly automated morphometric analysis [27–31]. For large whales, deep-learning approaches have been used to automatically detect individuals and extract body length, body width, and body condition from UAV imagery [30,31], while UAV photogrammetry has also been extended to small cetaceans, including body-length estimation in free-ranging finless porpoises [32,33]. Nevertheless, an integrated workflow that links large-format target discovery, orientation-aware body-surface segmentation, and differentiated morphometric extraction for both complete- and partial-body observations remains limited for small cetaceans. This limitation is particularly relevant to the Yangtze finless porpoise: because individuals are small and arbitrarily oriented, surfacing postures vary rapidly, body exposure is often incomplete, and body–water boundaries may be weak. An automated workflow adapted to these conditions is therefore needed to convert detected porpoises into reliable body-surface contours and image-derived morphometric parameters.

An oriented detection-guided SAM2 morphometric workflow, hereafter referred to as ODG-SAM2-Morph, is developed and evaluated for the Yangtze finless porpoise in the Yangtze River–Poyang Lake system. ODG-SAM2-Morph is defined as an integrated detection–segmentation–measurement workflow that uses YOLO26-OBb to localize small porpoise targets, applies SAM2 to extract prompt-guided visible body-surface masks, and performs sample-specific morphometric extraction for complete-body and partial-body samples. By linking oriented target localization, body-surface segmentation, and differentiated morphometric measurement, the workflow converts UAV observations into measurable image-derived morphometric parameters. The specific objectives are to: (1) detect small porpoise targets and obtain body-axis orientation from large-format UAV images; (2) extract visible body-surface masks using detection-guided SAM2 segmentation; (3) establish differentiated measurement strategies for complete-body and partial-body samples; (4) evaluate the reliability of detection, segmentation, and morphometric parameter extraction under complex natural aquatic conditions.

2. Materials

2.1. Study Area

This study was conducted in the Yangtze River–Poyang Lake system from 11 May to 13 May 2025. The natural distribution of the Yangtze finless porpoise (*Neophocaena asiaeorientalis*) is mainly concentrated in the middle and lower reaches of the Yangtze River mainstream and its connected lakes, among which Poyang Lake and Dongting Lake are the two major lake habitats of this species [34,35]. The Yangtze River–Poyang Lake system encompasses multiple aquatic environments, including the Yangtze River mainstream, tributary waters, river–lake transition zones, and open-lake areas. The aquatic environments differ markedly in hydrological connectivity, shoreline constraints, and water-surface backgrounds. The environmental differences enabled evaluation of UAV imagery-based target detection, body-surface contour extraction, and morphometric parameter calculation under different natural conditions. The study selected three survey subareas—Anqing, Duchang, and Kangshan—for UAV image acquisition to cover representative porpoise activity areas within the Yangtze River–Poyang Lake system (Figure 1). Anqing represents the mainstem waters of the Yangtze River, Duchang represents the open waters of Poyang Lake, and Kangshan represents a river–lake transition zone. The three survey subareas are located in Anqing (30.45°–30.55° N, 116.95°–117.05° E), Duchang (29.30°–29.45° N, 116.15°–116.30° E), and Kangshan (29.20°–29.30° N, 116.35°–116.50° E), respectively.

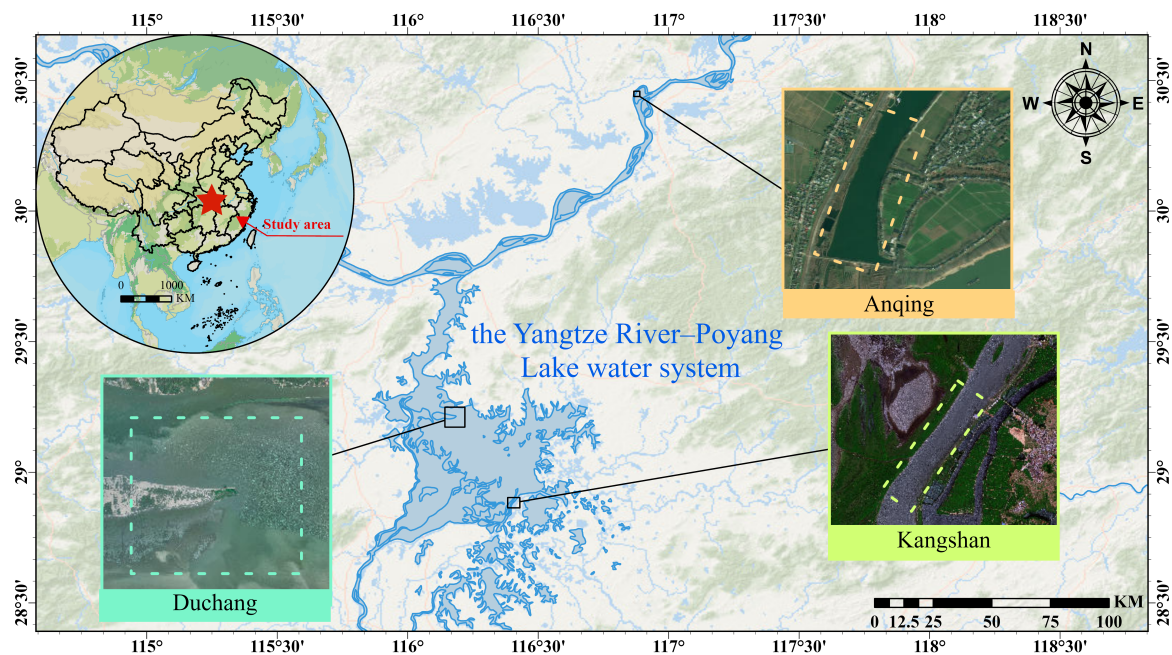


Figure 1. Location of the study area and the three representative survey subregions.

2.2. UAV Image Acquisition

UAV imagery of Yangtze finless porpoises was acquired using a fixed-wing CW-15 UAV (JOUAV, Chengdu, China) equipped with a CA-103 digital aerial camera (JOUAV, Chengdu, China), as shown in Figure 2. The CW-15 UAV had relatively long endurance, stable flight performance, and strong suitability for continuous aerial surveys over large aquatic areas, which made it capable of meeting the high-frequency observation requirements for the short surfacing behavior of porpoises in natural waters. High-resolution visible-light images were acquired using a CA-103 digital aerial camera equipped with a full-frame sensor, an effective resolution of 61 MP, an image size of 9504×6336 pixels, a pixel size of approximately $3.76 \mu\text{m}$, and a shutter speed range from $1/8000$ s to 30 s. The high pixel resolution and small pixel size enhanced the representation of water-surface details and provided a clear image basis for small-target porpoise detection, contour segmentation, and subsequent body-surface morphometric measurements.

During aerial surveys, the UAV was operated at an altitude of approximately 150 m above the water surface, with a cruising speed of about 16 m/s, an image acquisition interval of approximately 1 s, and a forward overlap of 80–90% between adjacent images. The flight and imaging parameter configuration ensured sufficient image resolution for porpoise targets while also balancing survey coverage, operational efficiency, and the need to minimize disturbance to porpoise activities. The short imaging interval and high forward overlap increased the temporal sampling density of consecutive images, helping to capture the brief surfacing process of porpoises and reduce the probability of missed observations. Each image was synchronously recorded with POS data, including position and attitude information. The POS records were used to evaluate image geometric status, perform single-frame geometric/tilt correction, and convert pixel-scale measurements into real-world dimensions. After each flight, the raw POS data were post-processed using JoPPS v1.0 software through a Post-Processed Kinematic (PPK) workflow to obtain more accurate GPS information at the camera exposure moments. PPK-refined POS data helped reduce the influence of UAV positioning errors on image geometry and pixel-to-real world scale conversion, thereby reducing scale-conversion uncertainty in subsequent morphometric measurements of Yangtze finless porpoises. The details of UAV survey flights and image acquisition parameters across the three study regions are summarized in Table 1.

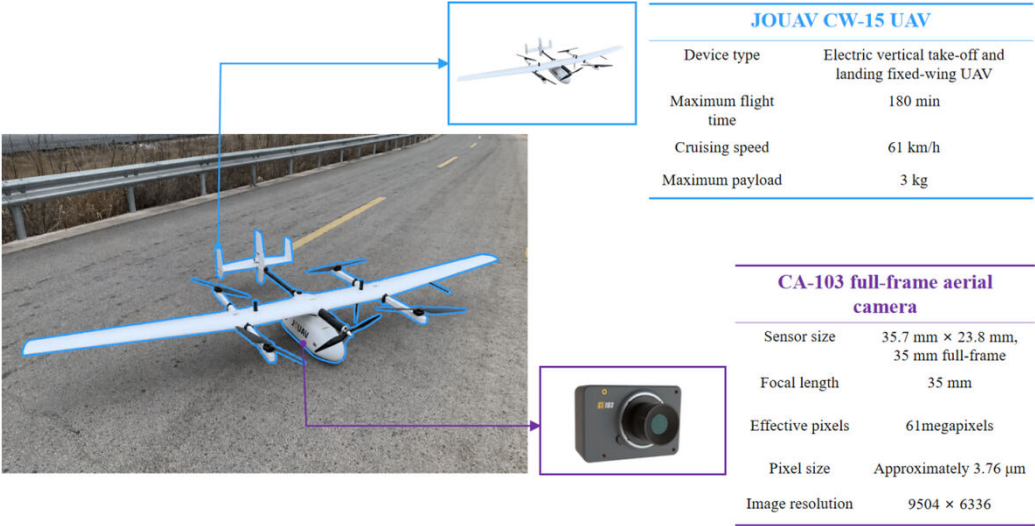


Figure 2. UAV platform and payload system used for high-resolution imagery acquisition. The fixed-wing CW-15 UAV is equipped with a CA-103 digital aerial camera for visible-light imaging.

Table 1. Summary of UAV survey flights and image acquisition across the study regions.

Water System	Region	Number of Flights	Time of Flight	Number of Photos
Poyang Lake	Duchang	3	170 min	12,228
Poyang Lake	Kangshan	2	106 min	6384
Yangtze Mainstream	Anqing	1	74 min	4706

2.3. Dataset Construction

Two datasets were constructed from the original UAV imagery: an oriented object detection dataset and a segmentation evaluation dataset, as illustrated in Figure 3. Oriented bounding boxes were manually annotated for the detection samples, while pixel-level body-surface masks were annotated for the segmentation samples using X-AnythingLabel v4.0.0. All annotations were reviewed before model evaluation. The object detection dataset was constructed through image screening, image tiling, hard-negative mining, group-constrained partitioning, and training-only data augmentation. A total of 23,318 original UAV images with a resolution of 9504 × 6336 pixels were acquired during the survey missions. After images dominated by terrestrial regions or lacking valid water-surface coverage were excluded, 329 source images containing visible Yangtze finless porpoises were identified, comprising 446 annotated porpoise instances. The 329 porpoise-containing images accounted for approximately 1.41% of the full image collection. Individual porpoises occupied only a small proportion of the original images. The positive source images were therefore cropped into 1024 × 1024-pixel tiles, and the tiling procedure retained only tiles containing clearly identifiable porpoise targets and sufficient surrounding water-surface context. Some tiles contained more than one porpoise, resulting in 446 annotated instances distributed across 391 positive tiles. The remaining water-surface images contained extensive homogeneous and repetitive background regions. Retaining all target-free tiles would have produced severe class imbalance and caused the dataset to be dominated by low-information backgrounds. Therefore, a preliminary YOLO detector was trained using the initial positive samples and applied to the remaining images with valid water-surface coverage to identify candidate hard-negative samples. The resulting false-positive detections were manually reviewed, and 257 representative hard-negative tiles containing major confusing objects or conditions, including non-target fish, sun glint, floating algae,

and birds, were retained. The preliminary detector was used only for hard-negative discovery and was not included in the final model evaluation. The final detection dataset therefore comprised 648 tiles, including 391 positive tiles and 257 hard-negative tiles. Temporally adjacent images could record the same surfacing event or contain highly similar water-surface backgrounds. Ignoring such temporal and hierarchical dependencies during dataset partitioning may introduce information leakage and result in overly optimistic estimates of model performance [36]. Positive tiles were therefore grouped according to field sighting records, acquisition timestamps, and the continuity of the original image identifiers. Tiles from the same surfacing event or continuous observation sequence were assigned to a common group, and multiple tiles cropped from the same source image were always retained together. The 391 positive tiles were ultimately organized into 121 distinct porpoise-sighting event groups. Hard-negative tiles were similarly grouped according to their source images and continuous flight sequences. Five-fold grouped cross-validation was then conducted at the event-group and source-sequence levels. In each run, three folds were used for training, one for validation, and one for testing, with the fold roles rotated across five runs. Mosaic, MixUp, random rotation, scaling, translation, HSV adjustment, and horizontal flipping were applied only to the training folds [37,38].

The segmentation evaluation dataset consisted of 121 porpoise-positive images selected from the sighting records, including seven independent complete-body samples and 114 partial-body samples. Complete-body samples were defined as images in which the rostrum, caudal-peduncle end, and a relatively continuous body contour could be identified, with an unambiguous body-axis direction. Partial-body samples included images in which either body endpoint was not visible or the body axis was strongly curved, preventing reliable extraction of full body length and fixed-position body-width profiles. For segmentation evaluation, image clarity was used as the primary selection criterion, and temporally adjacent frames could be retained when they represented different surfacing postures. For morphometric agreement analysis, however, only one representative frame was selected from each sighting record, prioritizing trunk exposure, contour clarity, and an unambiguous body-axis direction. All pixel-level body-surface masks were manually annotated and subsequently verified. The dataset was used only to compare SAM2 prompt strategies and model sizes and was not included in SAM2 training or fine-tuning. The manually annotated masks also served as reference data for evaluating the extraction of body-surface morphometric parameters.

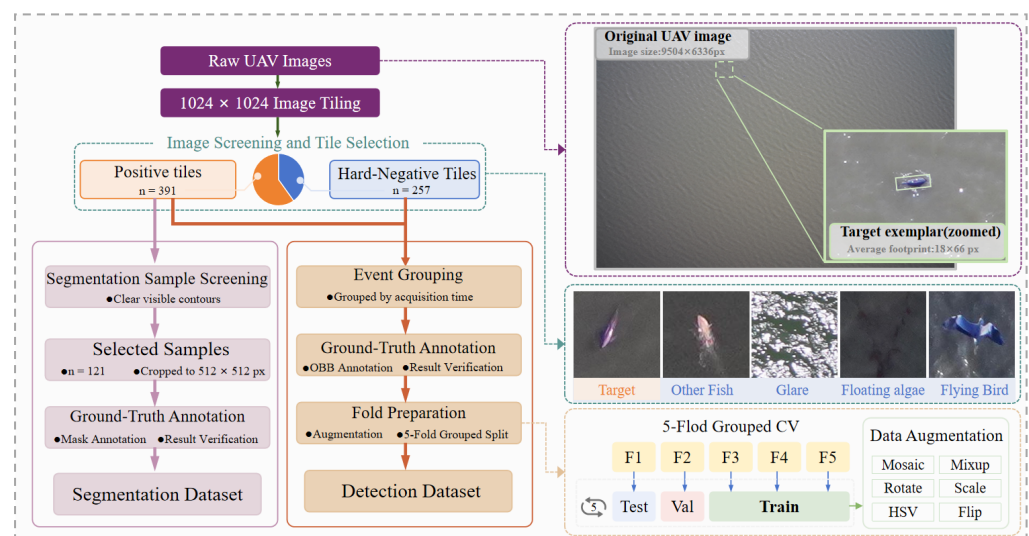


Figure 3. Event-grouped dataset construction and augmentation workflow for UAV-based porpoise detection.

3. Methods

3.1. Morphometric Information Extraction

Morphometric information extraction was performed using geometrically corrected UAV imagery and the corresponding binary masks of porpoise body surfaces. The natural surfacing process of porpoises is highly transient, and individuals may appear in different behavioral states, including breathing, swimming, turning, and rolling. Consequently, the visible body-surface extent, contour completeness, and continuity of the body axis varied substantially among images. The analysis first classified samples as complete-body or partial-body according to visible-contour completeness and then applied a different parameter extraction strategy to each category. The analysis preferentially selected the frame with the largest visible body-surface extent and clearest contour from consecutive images of the same surfacing record (Figure 4).

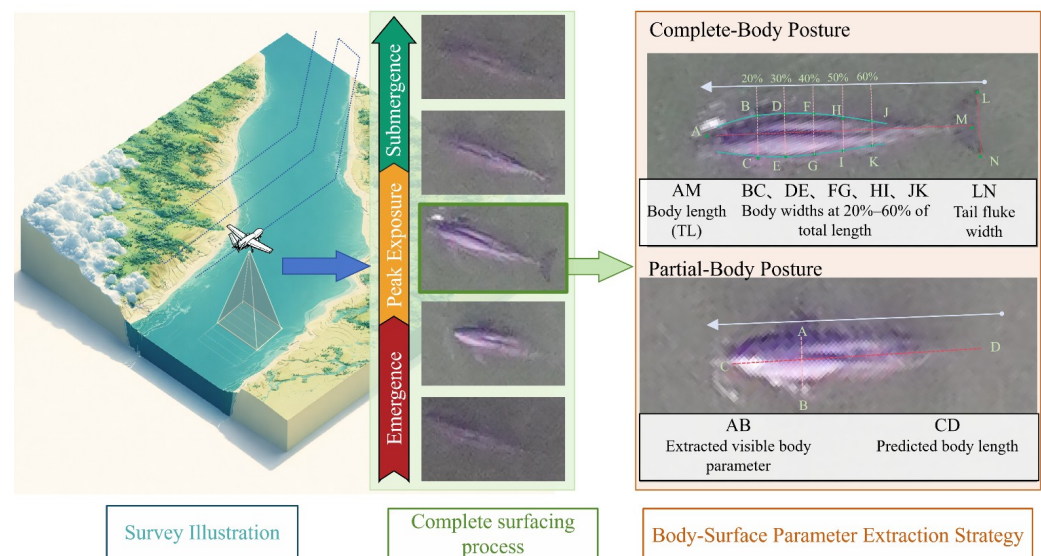


Figure 4. Schematic illustration of the UAV survey and body-surface morphometric parameter extraction.

Complete-body samples were defined as images in which the rostrum, caudal-peduncle end, and a relatively continuous body contour could be identified, with an unambiguous body-axis direction. The body axis of each complete-body sample was first determined from the body-surface mask. The projection range of the mask along the body axis was then used to locate the rostrum and caudal-peduncle end, from which body length was calculated. Width scanning was subsequently performed perpendicular to the body axis to obtain body-width parameters at different relative positions. $W_{\max}$ was defined as the maximum body width perpendicular to the body axis in complete-body samples. W_{20} , W_{30} , W_{40} , W_{50} , and W_{60} represented body widths measured at 20%, 30%, 40%, 50%, and 60% of body length from the rostrum, respectively. The fixed-position width parameters described longitudinal variation in body-surface width along the main body axis.

Partial-body samples were defined as images in which the rostrum or caudal-peduncle end was absent, or in which the body axis was strongly curved, making it difficult to reliably obtain full body length and fixed-position width profiles. For partial-body samples, only the maximum visible body width, denoted as $W_{\max, \text{vis}}$, was extracted. $W_{\max, \text{vis}}$ was defined as the maximum transverse width of the body-surface region visible in each image, measured perpendicular to the local body axis. In practice, the visible body-surface mask was first extracted from the SAM2 segmentation result. The local body axis was then determined based on the principal axis of the mask or the local contour direction. Width scanning was

subsequently performed perpendicular to the local body axis, and the maximum width within the visible region was recorded as $W_{\max, \text{vis}}$.

For exploratory purposes, a potential body-length estimate for partial-body samples was derived using the preliminary relationship between maximum body width and body length observed in the seven complete-body samples. The ratio of maximum body width to body length was calculated as follows:

$$k = \frac{W_{\max}}{TL} \quad (1)$$

where k represents the maximum body width-to-body length ratio derived from complete-body samples, TL is the measured body length, and $W_{\max}$ is the maximum body width of complete-body samples. For partial-body samples, the estimated body length was then calculated as:

$$TL_{\text{est}} = \frac{W_{\max, \text{vis}}}{k} \quad (2)$$

where TL_{est} is the potential body-length estimate of the partial-body sample, and $W_{\max, \text{vis}}$ is the maximum visible body width under partial surfacing or curved-posture conditions. The basic assumption of this method is that when the visible region of a partial-body sample includes the main transverse scale of the trunk, its maximum visible body width can approximate the individual maximum body width to some extent. Samples with an excessively small visible region, unclear body-axis direction, or severe contour interference caused by splashes or surface reflection were excluded from body-length estimation to avoid bias caused by low-quality samples. Finally, pixel-scale measurements were converted into real-world morphometric parameters according to the spatial resolution of the UAV imagery. The real value of a linear parameter was calculated as:

$$L_{\text{real}} = L_{\text{px}} \times \text{GSD} \quad (3)$$

where L_{real} represents the real-world length, L_{px} represents the measured pixel length, and GSD represents the ground sampling distance of the UAV image. Through this procedure, a continuous workflow was established from YOLO26-OBB-based target localization and SAM2-based body-surface segmentation to porpoise morphometric parameter calculation. The differentiated measurement strategies for complete-body and partial-body samples improved the utilization efficiency of UAV image samples while maintaining the reliability of the extracted parameters.

3.2. Image Processing

3.2.1. YOLO26-OBB-Based Candidate Localization

In ODG-SAM2-Morph, YOLO26-OBB served as the front-end target localization and prompt-generation model [39]. The model detected Yangtze finless porpoises from 1024×1024 -pixel UAV image tiles and predicted oriented bounding boxes containing target position, spatial extent, and rotation angle. Each oriented bounding box defined a candidate region of interest and provided an initial estimate of body-axis orientation for subsequent SAM2 segmentation. The detection boxes were not used as final morphometric boundaries because rectangular regions could not represent visible body contours at the pixel level.

The YOLO26-OBB architecture comprised a Backbone, Neck, and oriented detection Head, as illustrated in Figure 5. The Backbone extracted multi-scale features using Conv, C3K2, SPPF, and C2PSA modules. The Neck fused features from different spatial resolutions, while the oriented detection Head predicted the center coordinates, dimensions, rotation angle, confidence score, and class probability of each target. The resulting

multi-scale representation supported porpoise detection under variations in target size, orientation, and water-surface background. The elongated shape and variable orientation of surfaced porpoises motivated the use of oriented detection. The long axis of each predicted box provided a directional prior for ROI rotation, and the tighter spatial coverage reduced the surrounding water included in the candidate region. The ROI was cropped and rotated according to the predicted angle so that the porpoise body axis was approximately aligned with the horizontal image axis. The transformed box then served as the spatial and directional prompt for SAM2-based body-surface segmentation. Porpoise localization was formulated as an oriented bounding box detection task, in which an oriented bounding box (OBB) was used to describe the position, scale, and orientation of each porpoise target. The use of OBB detection was based on two main considerations. First, the long-axis direction of the rotated bounding box can effectively reflect the body-axis orientation of a porpoise in the image, providing an initial geometric reference for subsequent body-length direction determination, body-width profile construction, and morphometric parameter calculation. This is particularly important when porpoises appear at oblique angles, change direction, or are not horizontally aligned in the image, as the orientation information provided by OBBs can reduce measurement bias caused by relying solely on the image coordinate axes. Second, the rotated bounding box can more tightly enclose the main body of the porpoise target. The spatial consistency between the region inside the bounding box and the target mask can serve as an important reference for evaluating the quality of subsequent SAM2 segmentation results, thereby providing geometric constraints for mask validity assessment and body-surface parameter extraction.

Model training and inference were conducted on an NVIDIA RTX 2080 Ti GPU (NVIDIA Corporation, Santa Clara, CA, USA) with 11 GB memory. The experimental environment consisted of Python 3.10.19, PyTorch 2.5.0, and CUDA 12.4. The input image size was set to 1024×1024 pixels. The detector was trained for 300 epochs with a batch size of 16. Stochastic gradient descent was used as the optimizer, with an initial learning rate of 0.01 and a final learning-rate factor of 0.01. Momentum and weight decay were set to 0.937 and 0.0005, respectively.

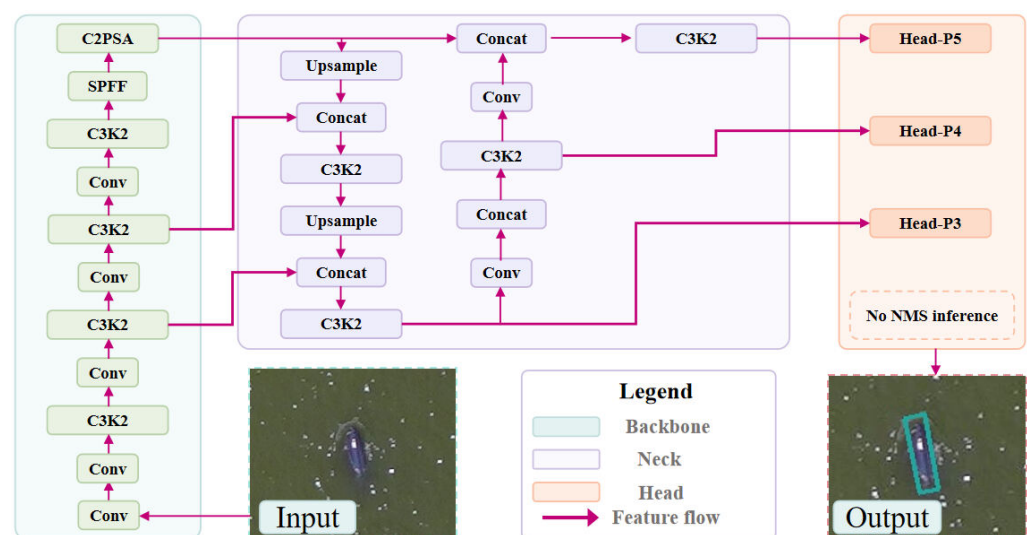


Figure 5. Network architecture of the YOLO26-OBB model for Yangtze finless porpoise detection in UAV imagery.

3.2.2. SAM2

SAM2 is a promptable segmentation model that generates pixel-level masks of target regions from an input image and associated prompts (Figure 6). SAM2 was used to

extract pixel-level body-surface masks from the candidate regions detected by YOLO26. Considering the small target size, weak body-surface boundaries, and complex background interference in UAV imagery of Yangtze finless porpoises, box and point prompts were mainly used to construct the SAM2 inputs. For each detection, a 512×512 -pixel region of interest (ROI) was cropped around the center of the oriented bounding box. The ROI was then rotated according to the predicted orientation angle so that the long axis of the porpoise was approximately aligned with the horizontal image axis. The four vertices of the oriented bounding box were transformed using the same rotation, and their minimum enclosing axis-aligned rectangle was used as the orientation-normalized box prompt. After segmentation, the predicted mask was mapped back to the original ROI coordinates using the inverse rotation transformation.

Three representative prompt strategies were evaluated to progressively examine the contribution of spatial and orientation information to SAM2 porpoise segmentation: None, Box + 1FG, and R-Box + 1FG. The None condition served as the reference configuration without an explicit spatial prompt derived from detection. Box + 1FG combined a horizontal bounding box with one foreground point to provide a coarse spatial constraint and explicit target identification. R-Box + 1FG replaced the horizontal box with a bounding box aligned with the porpoise body axis, thereby providing spatial orientation information while retaining the same foreground point constraint. For the orientation ablation, the horizontal Box was generated as the minimum axis-aligned enclosing rectangle of the same predicted OBB. Therefore, target localization was kept fixed while only the orientation information was removed. This strategy was designed to reduce the excessive inclusion of surrounding water that can occur when horizontal boxes are applied to elongated porpoises appearing at arbitrary orientations in UAV images. All prompts were generated geometrically from the detection results and ROI coordinates without reference to the manually annotated masks. This progressive design enabled the effects of target localization and body-axis orientation on segmentation performance to be evaluated under otherwise consistent prompting conditions.

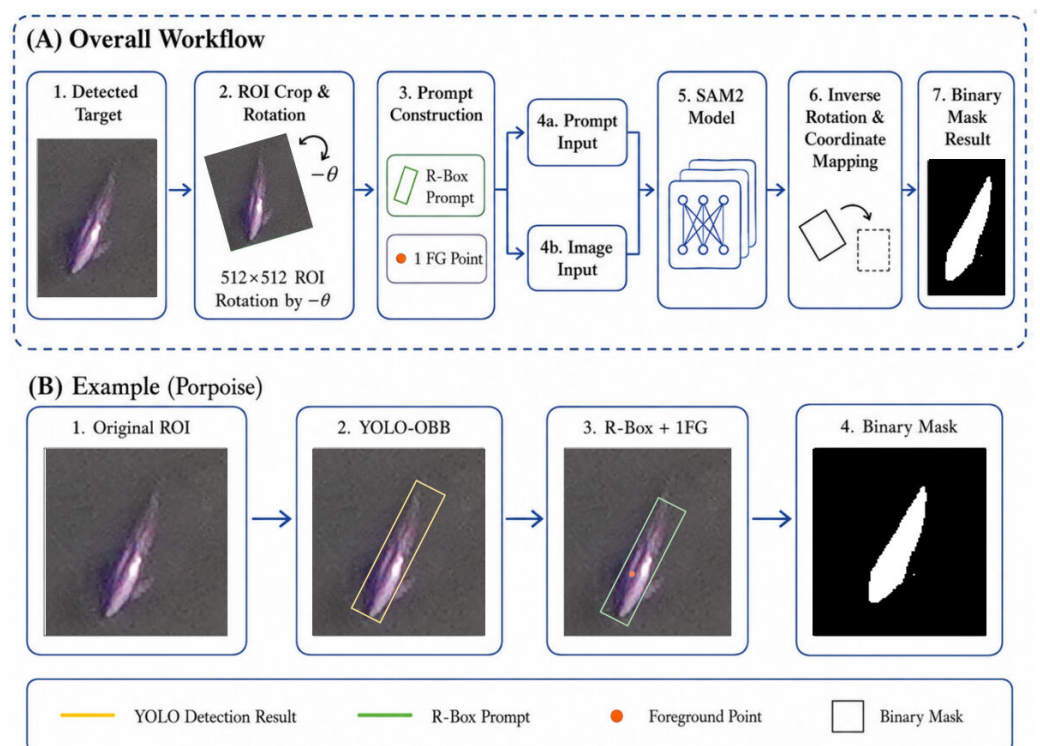


Figure 6. ROI extraction and prompt construction for SAM2-based porpoise body-surface segmentation.

After SAM2 inference, the generated body-surface masks were post-processed to obtain binary masks suitable for morphometric parameter extraction. The post-processing procedure retained the target mask that was spatially consistent with the prompt region and removed small noisy regions. The resulting binary masks of porpoise body surfaces were then used as the basis for extracting morphometric parameters, including body length, body width, maximum visible body width, and body-width profiles.

3.3. Overall Workflow

To achieve non-contact and automated extraction of body-surface morphometric parameters of Yangtze finless porpoises from UAV imagery, a coordinated detection–segmentation–measurement framework was developed based on YOLO26-OBB and SAM2. The overall technical workflow is shown in Figure 7. Within the overall framework, YOLO26-OBB was used as the front-end target localization module to automatically retrieve and locate Yangtze finless porpoises from large-format UAV images of aquatic environments. Given the small target size and complex water-surface backgrounds, YOLO26-OBB was first used to screen candidate porpoise targets and generate oriented bounding boxes containing target position, scale, and orientation information. These oriented bounding boxes effectively represented the elongated body shape and body-axis direction of porpoises and were further converted into spatial prompts to guide SAM2 in performing pixel-level segmentation within local candidate regions. Through this collaborative strategy, unconstrained segmentation by SAM2 over the entire image was avoided, while its lack of active target detection capability was also compensated for. In the framework, YOLO26-OBB was responsible for target discovery and spatial constraint generation in large-area images, whereas SAM2 focused on fine boundary extraction within local regions. Therefore, the detection boxes were not used as the final measurement boundaries, but served as spatial priors and geometric prompts for the segmentation model. Finally, the binary masks generated by SAM2 provided the pixel-level basis for calculating body length, body width, and body-width profiles, enabling a continuous transition from object-level localization to measurable body-surface contour extraction.

The feasibility of the UAV imagery-based framework was evaluated from three aspects. First, the performance of YOLO26-OBB and SAM2 was assessed for porpoise localization and body-surface segmentation, with attention to the effects of prompt strategies and model sizes. Second, errors in morphometric parameter extraction were analyzed by comparing automatic and manual measurements, considering factors such as shadows, water-surface reflections, ripples, splashes, and surfacing posture. Third, the applicability of complete-body and partial-body samples for morphometric analysis was examined, particularly with respect to the stability of maximum visible body width and its potential use in body-length inference.

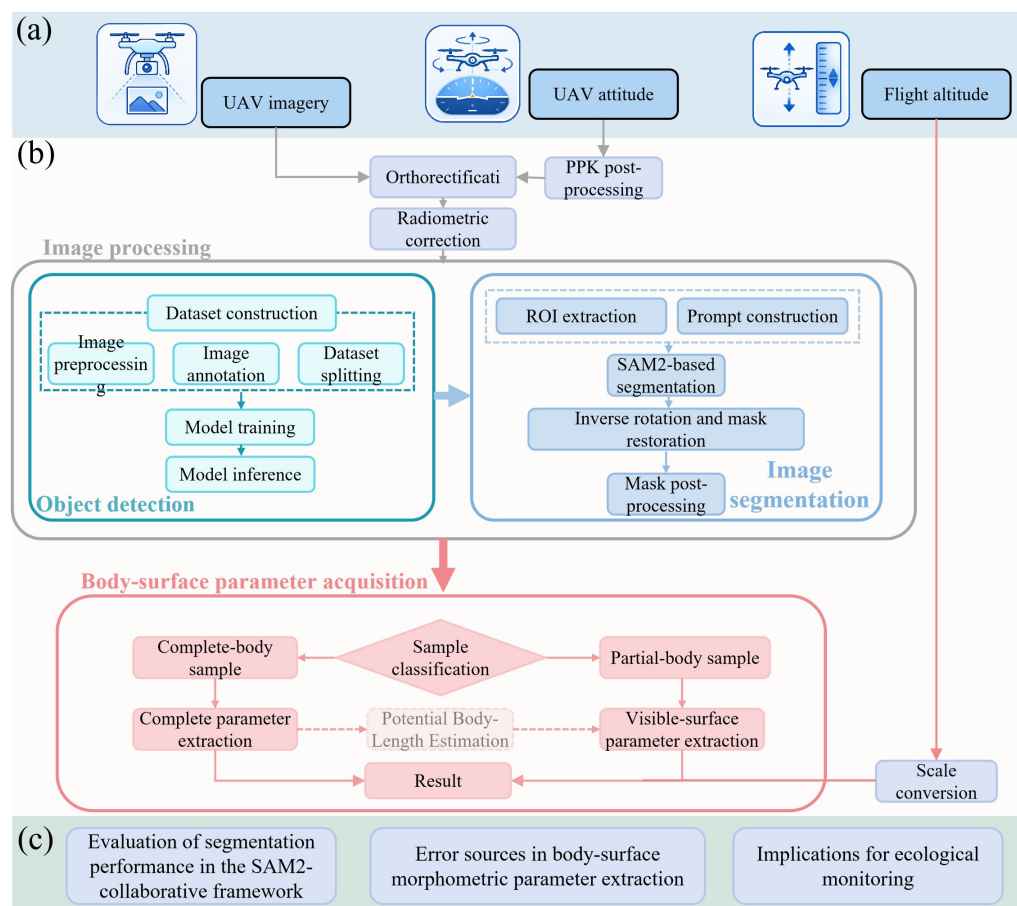


Figure 7. Overall workflow for UAV-based detection, segmentation, and morphometric parameter estimation of the Yangtze finless porpoise: (a) UAV imagery acquisition and associated flight information; (b) image processing, object detection, SAM2-based segmentation, and body-surface morphometric parameter extraction; (c) evaluation of segmentation performance, error sources, and implications for ecological monitoring.

3.4. Evaluation Metrics

The framework was evaluated at three levels: oriented target detection, body-surface segmentation, and morphometric parameter extraction. Detection metrics were calculated for each test fold and summarized across the five grouped cross-validation runs. Segmentation and morphometric measurements were evaluated against manually annotated masks and manual measurements, respectively. Detection performance was assessed using Precision, Recall, mAP@0.5, and mAP@0.5:0.95:

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively. The mAP@0.5 metric was calculated at an intersection-over-union threshold of 0.5, whereas mAP@0.5:0.95 was averaged over IoU thresholds from 0.5 to 0.95 at intervals of 0.05. Detection results from the five test folds were reported as mean $\pm$ standard deviation. Segmentation performance was evaluated by comparing the SAM2-predicted masks with

manually annotated body-surface masks. Pixel-level Precision, Recall, intersection over union, and Dice coefficient were calculated as:

$$P_{\text{seg}} = \frac{|M_p \cap M_g|}{|M_p|} \quad (6)$$

$$R_{\text{seg}} = \frac{|M_p \cap M_g|}{|M_g|} \quad (7)$$

$$\text{IoU} = \frac{|M_p \cap M_g|}{|M_p \cup M_g|} \quad (8)$$

$$\text{Dice} = \frac{2|M_p \cap M_g|}{|M_p| + |M_g|} \quad (9)$$

where M_p and M_g represent the predicted and manually annotated masks, respectively. Precision quantified the proportion of predicted body-surface pixels that matched the reference mask, while Recall quantified the proportion of reference pixels retained by the predicted mask. IoU and Dice were used to assess overall spatial agreement between the two masks. Morphometric accuracy was evaluated by comparing automated and manual measurements of body length, tail fluke width, fixed-position body widths, maximum body width, and maximum visible body width. Pixel measurements were converted to centimeters using the corresponding image GSD before statistical analysis. Measurement error was quantified using the mean absolute error (MAE) and mean absolute percentage error (MAPE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i| \quad (10)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right| \times 100\% \quad (11)$$

where Y_i denotes the manual measurement, $\hat{Y}_i$ denotes the automated measurement, and n denotes the number of observations. The systematic difference between the automated and manual measurements was assessed using the Bland–Altman means bias. The paired measurement difference was defined as:

$$d_i = \hat{Y}_i - Y_i \quad (12)$$

and the mean bias was calculated as:

$$\hat{d} = \frac{1}{n} \sum_{i=1}^n d_i \quad (13)$$

A positive mean bias indicated an overall overestimation by the automated method, whereas a negative value indicated an overall underestimation. Measurement agreement was further evaluated using a two-way mixed-effects, absolute-agreement, single-measure intraclass correlation coefficient, denoted as ICC(A,1), and its 95% confidence interval was reported. MAE and MAPE quantified the magnitude of measurement error, the Bland–Altman means bias described the average systematic difference, and the ICC confidence interval characterized the uncertainty in measurement agreement.

4. Results

4.1. Image Processing Results

The proposed workflow was evaluated in terms of oriented porpoise detection and body-surface image segmentation. Despite the limited dataset size of 648 tiles, online data augmentation substantially improved YOLO26-OBB detection performance. Without augmentation, Precision, Recall, mAP@0.5, and mAP@0.5:0.95 were 0.7078, 0.6224, 0.6836, and 0.3604, respectively. Online augmentation increased the corresponding values to 0.9091, 0.8197, 0.9155, and 0.5513. The mean best epoch also increased from 116.0 to 283.8, indicating that augmentation supported a longer training process before convergence. Complete fold-level results are reported in Table 2.

Table 2. Complete five-fold cross-validation results for YOLO26n-OBB.

CV Fold	Test-Set Images	Optimal Epoch	Precision	Recall	mAP@0.5	mAP@0.5:0.95
0	129	295	0.927	0.844	0.925	0.578
1	130	288	0.942	0.766	0.898	0.541
2	129	268	0.891	0.860	0.936	0.513
3	130	283	0.934	0.876	0.950	0.592
4	130	285	0.852	0.753	0.868	0.532
Mean ± SD	129.6	283.8 ± 9.9	0.909 ± 0.0374	0.820 ± 0.0565	0.916 ± 0.0329	0.551 ± 0.0329

The consistent improvements across all detection metrics demonstrate the effectiveness of the selected augmentation strategies for training on the small dataset. Nevertheless, mAP@0.5:0.95 remained substantially lower than mAP@0.5, indicating reduced localization accuracy under stricter IoU thresholds. The predicted oriented boxes were therefore suitable for locating porpoise targets and estimating body-axis orientation, but could not reliably represent the fine body-surface boundaries required for morphometric measurement. For body-surface segmentation, SAM2 was applied to the candidate regions localized by YOLO26-OBB. Among the evaluated configurations, the small SAM2 model using the R-Box + 1FG prompt achieved the best overall performance, with a mean IoU of 0.824, a mean Dice coefficient of 0.901, a Precision of 0.883, and a Recall of 0.932. The relatively high Recall indicates that most manually annotated visible body-surface regions were retained in the predicted masks. The mean IoU and mean Dice coefficient also indicate strong spatial agreement between the predicted and reference masks. Precision was slightly lower than Recall, suggesting that adjacent shadows, water-surface reflections, or low-contrast transition regions were occasionally included in the predicted masks. Overall, SAM2 produced pixel-level body-surface contours that were more suitable for morphometric parameter extraction than the oriented detection boxes. Figure 8 illustrates the complete image-processing sequence from original UAV imagery to porpoise detection and body-surface segmentation. YOLO26-OBB identified small porpoise targets from large water-surface images and provided their positions and approximate body-axis orientations. SAM2 subsequently performed pixel-level segmentation within the detected regions and generated binary masks of the visible body surface. The representative samples demonstrate the complementary roles of the two stages: oriented detection provided target localization and directional constraints, while image segmentation supplied the detailed contours required for subsequent morphometric measurements.

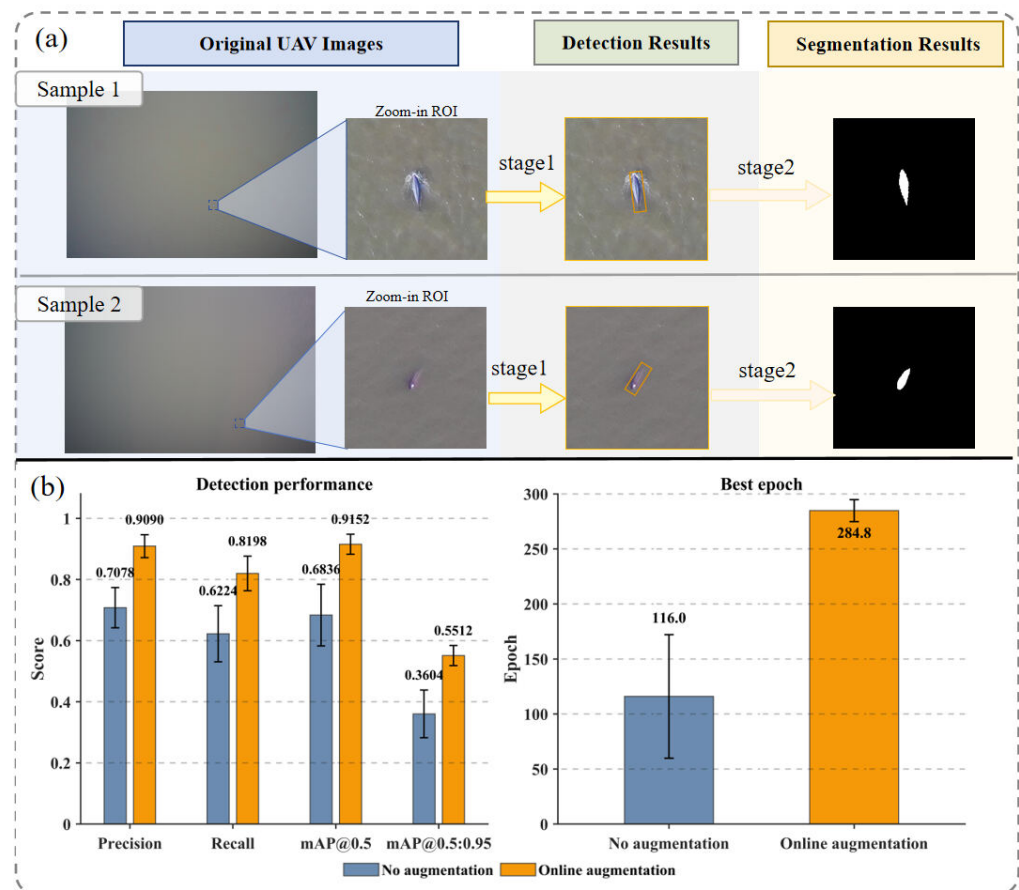


Figure 8. Representative image processing results from original UAV imagery to YOLO26-OBb detection and SAM2 segmentation: (a) image processing workflow from original UAV images to detection results and segmentation results; (b) comparison of detection performance and best training epoch with and without online augmentation.

4.2. Estimation Results of Finless Porpoise Body-Surface Parameters

Body-surface morphometric parameters were extracted from the SAM2-derived masks and compared with manual measurements. At the survey altitude of approximately 150 m, the UAV imagery had a ground sampling distance of approximately 1.61 cm/pixel. Pixel-based measurements were therefore converted to centimeters using the corresponding image GSD before statistical analysis. Complete-body and partial-body samples were evaluated separately because body exposure and contour completeness varied among surfacing images. Seven complete-body samples with identifiable rostra, caudal-peduncle ends, and continuous body contours were used to evaluate body length, tail fluke width, widths at 20%, 30%, 40%, 50%, and 60% of body length, and maximum body width. As shown in Figure 9a, the automatically extracted contours generally followed the manually delineated body boundaries. Across all complete-body parameters, the MAE was 3.25 cm, the MAPE was 14.28%, and the Bland–Altman mean bias was 2.67 cm. The 95% confidence interval of the ICC ranged from 0.982 to 0.997, indicating high agreement between automated and manual measurements. Parameter-specific MAPEs were 1.90% for body length, 7.90% for tail fluke width, 14.8%, 14.2%, 15.5%, 16.9%, and 24.3% for widths at 20–60% of body length, respectively, and 18.7% for maximum body width. Body length showed the lowest error, whereas transverse width measurements were more sensitive to local boundary displacement. Sample S2 showed the largest deviations across several width-related parameters, mainly because of uncertainty along the visible body boundary.

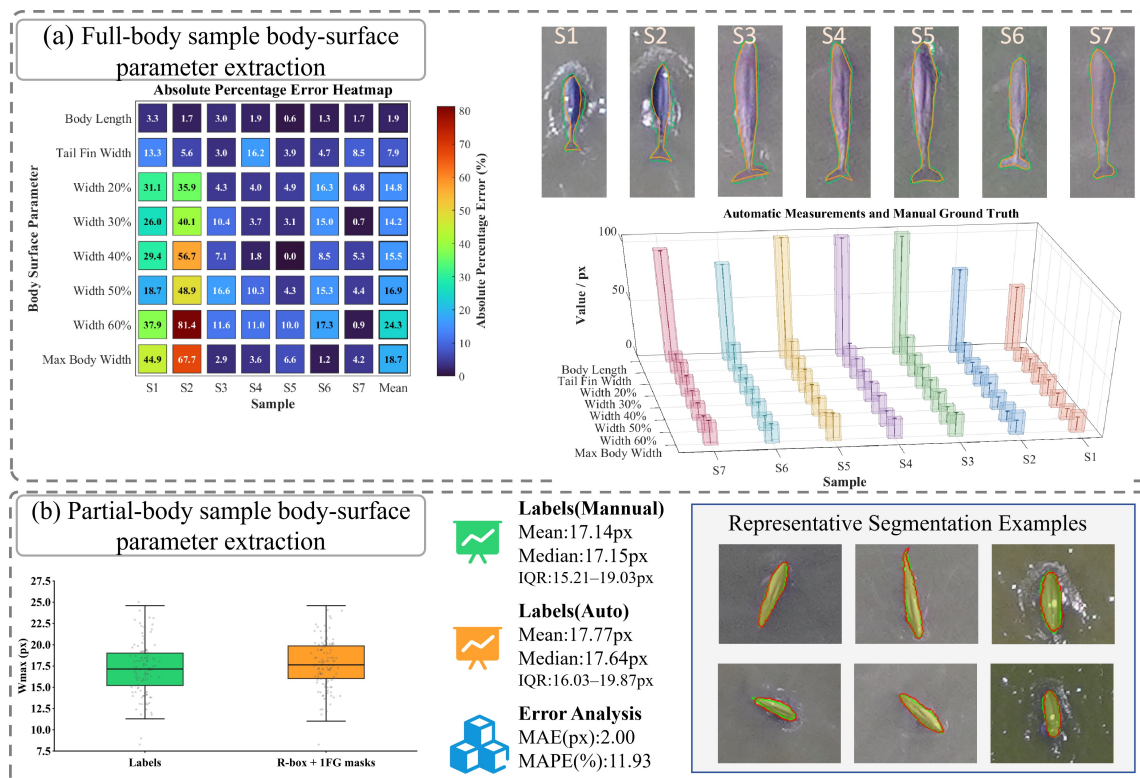


Figure 9. Morphometric parameter extraction results for complete-body and partial-body samples. (a) Complete-body sample body-surface parameter extraction; (b) partial-body sample body-surface parameter extraction. The green contours indicate the automatically extracted body-surface boundaries, whereas the yellow contours indicate the manually annotated ground-truth boundaries.

Maximum visible body width was evaluated using 114 partial-body samples. Manual measurements had a mean of 28.61 cm, a median of 28.53 cm, and an interquartile range of 25.64–31.51 cm. Using the mean maximum body width-to-body length ratio derived from the complete-body samples ($k = 0.197$), the automated mean maximum visible body width corresponded to an estimated mean body length of approximately 150.0 cm. The MAE and MAPE were 3.32 cm and 11.93%, respectively. Bland–Altman analysis produced a mean bias of 0.94 cm, indicating slight overall overestimation by the automated method. The ICC 95% confidence interval ranged from 0.444 to 0.722. The broadly comparable central tendencies and distribution ranges suggest that the SAM2-derived masks supported the extraction of maximum visible body width from partial-body images. Larger deviations mainly occurred when trunk exposure was limited or when low contrast, water-surface reflections, and ripples reduced boundary clarity.

5. Discussion

5.1. Evaluation of Segmentation Performance in the SAM2-Collaborative Framework

The comparison of SAM2 configurations showed that segmentation performance was affected more strongly by prompt design than by model size. Without explicit spatial guidance, SAM2 showed limited ability to consistently distinguish the visible porpoise body surface from complex water-surface backgrounds, with a mean IoU of 0.371 and a mean Dice coefficient of 0.487. Introducing detection-derived prompts substantially improved segmentation accuracy. Among the three principal prompt configurations, the small SAM2 model with R-Box + 1FG achieved the best overall performance, reaching a mean IoU of 0.824, a mean Dice coefficient of 0.901, a Precision of 0.883, and a Recall of 0.932. For baseline comparison, Mask R-CNN and YOLO11n-Seg were evaluated on the same

segmentation dataset using grouped five-fold cross-validation and the same segmentation metrics. Mask R-CNN and YOLO11n-Seg achieved mean IoU values of 0.796 and 0.730 and mean Dice coefficients of 0.887 and 0.844, respectively. These baseline comparisons indicate that the detection-guided SAM2 strategy provided competitive or improved body-surface segmentation performance while retaining a prompt-based formulation without task-specific retraining of SAM2. Meanwhile, the relatively small differences among SAM2 model sizes under the same prompting condition further suggest that, for small porpoise targets with weak and partially submerged boundaries, the quality of spatial and directional guidance was more influential than simply increasing model scale.

The selection of YOLO26-OBB as the front-end detector was based on the combined requirements of detection accuracy, computational efficiency, and the quality of spatial guidance provided to SAM2. Although YOLO11n-OBB achieved the highest mAP@0.5 of 0.952, YOLO26n-OBB required fewer computational operations and provided the shortest end-to-end inference time, with 2.648 M parameters, 13.913 GFLOPs, and an inference time of 6.514 ms/image, compared with 8.095 ms/image for YOLO11n-OBB and 8.223 ms/image for YOLOv8n-OBB. This efficiency advantage is particularly relevant for UAV surveys involving large numbers of images that require target screening before segmentation and morphometric extraction. In addition to computational efficiency, the oriented-box formulation provided an important downstream advantage. A conventional horizontal Box can constrain the approximate target location but often includes substantial irrelevant water background when the elongated porpoise body appears obliquely or changes swimming direction. In contrast, the R-Box better conforms to the body-axis orientation and visible body extent, thereby providing SAM2 with a tighter spatial prior. The Box-versus-R-Box comparison supports this design choice: replacing the horizontal box with the rotated box increased IoU from 0.702 to 0.817 and Dice from 0.814 to 0.897. Taken together, these findings support the selection of YOLO26-OBB not because it achieved the highest standalone detection accuracy, but because it provided a favorable balance between detection capability, processing efficiency, and orientation-aware spatial guidance for downstream SAM2 segmentation.

An additional prompt-density experiment on the representative sample in Figure 10 showed that increasing the number of foreground and background points did not further improve segmentation performance. The key factor was not the number of prompts, but whether the prompts accurately described the target position, body-axis orientation, and visible body-surface region. The representative example in Figure 10 shows that R-Box alone already provided effective spatial and directional guidance, achieving a Dice of 0.913 and an IoU of 0.840. After adding one foreground point within the visible body region, the R-Box + 1FG strategy further improved the segmentation result, with Dice and IoU increasing to 0.931 and 0.871, respectively. However, adding more background or foreground points did not produce additional gains. When two or four background points were added, the Dice and IoU decreased to 0.912/0.838 and 0.902/0.821, respectively. A further increase to R-Box + 2FG + 8BG resulted in similar or slightly lower performance, with a Dice of 0.901 and an IoU of 0.820. These results indicate that, once the rotated box has provided a reliable directional constraint, a single foreground point is sufficient to confirm the target body region, whereas excessive prompt points may introduce uncertainty from water-surface ripples, reflections, and boundary transition zones. Therefore, R-Box + 1FG was selected as the most suitable prompt strategy for porpoise body-surface segmentation, as it balances directional constraint and target confirmation while avoiding over-prompting.

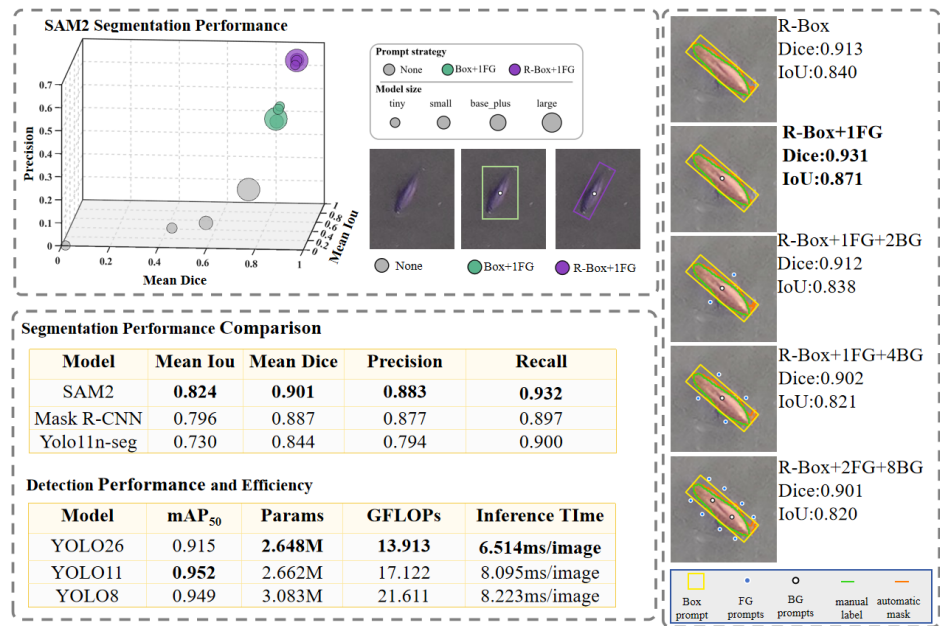


Figure 10. Effects of prompt strategy and SAM2 model size on porpoise body-surface segmentation performance.

5.2. Error Sources in Body-Surface Morphometric Parameter Extraction

Errors in body-surface morphometric parameter extraction were not caused solely by the segmentation model, but were jointly influenced by upstream detection uncertainty, UAV imaging conditions, water-surface interference, and porpoise surfacing posture [40]. In the proposed detector–segmenter workflow, errors in the OBB predicted by YOLO may alter the spatial prompt supplied to SAM2 and consequently affect downstream body-surface segmentation. Meanwhile, changes in illumination, water-surface reflection, ripples, and splashes can reduce the contrast between the exposed body surface and the surrounding water, thereby affecting mask-boundary delineation.

To quantify detection-to-segmentation error propagation, controlled translation, scale, and rotation perturbations were independently applied to the original YOLO-predicted OBB while the remaining geometric parameters were kept unchanged. Translation perturbations were defined as 2.5%, 5%, and 10% of the original OBB diagonal length and applied along the positive and negative directions of the major and minor axes. Scale perturbations were introduced by synchronously scaling the long and short sides by ±5%, ±10%, and ±20%, whereas rotation perturbations were generated by changing the OBB angle by ±2.5°, ±5°, and ±10°. These three levels represented mild, moderate, and severe perturbations, respectively. Only one error factor was changed in each experiment, and the results from positive and negative perturbation directions were averaged. All perturbation experiments were conducted using the R-Box + 1FG configuration.

As shown in Table 3, the detection-to-segmentation error propagation analysis revealed different sensitivities to the three types of OBB perturbations. Translation errors had the strongest downstream effect. Mild and moderate translation perturbations reduced mean IoU by 0.72% and 3.38%, respectively, whereas severe translation reduced mean IoU and mean Dice by 23.41% and 17.90%. Scale perturbations produced a smaller but progressively increasing effect, with severe perturbations reducing mean IoU and mean Dice by 7.65% and 4.58%, respectively. In contrast, SAM2 segmentation was comparatively robust to angular errors. Rotation perturbations up to 5° produced negligible changes in mean IoU and mean Dice, while a 10° perturbation reduced them by only 0.79% and 0.47%, respectively. The small positive changes observed under some rotation conditions were close to zero and

were therefore interpreted as model variation rather than actual performance improvement. These results indicate that positional displacement of the detected OBB has the greatest downstream influence because it directly shifts the spatial prior supplied to SAM2, whereas moderate angular deviations can be tolerated as long as the visible body remains sufficiently constrained by the R-Box. Because subsequent morphometric parameters are extracted from the segmented body contour, detection-induced segmentation deviations may be further transferred to body-length and body-width measurements.

Table 3. Relative changes in SAM2 segmentation performance under controlled perturbations of YOLO-predicted OBBs.

Perturbation	Severity	Mean IoU	Mean Dice	Precision	Recall
Translation	Mild	−0.72%	−0.42%	−0.43%	−0.35%
	Moderate	−3.38%	−2.05%	−1.85%	−2.01%
	Severe	−23.41%	−17.90%	−16.26%	−18.74%
Scale	Mild	−0.44%	−0.26%	−0.19%	−0.20%
	Moderate	−1.52%	−0.90%	−0.54%	−0.76%
	Severe	−7.65%	−4.58%	−2.87%	−3.88%
Rotation	Mild	+0.06%	+0.03%	+0.27%	−0.19%
	Moderate	+0.02%	+0.00%	+0.70%	−0.63%
	Severe	−0.79%	−0.47%	+1.95%	−2.61%

In addition to upstream detection errors, three major image-level factors contributed to morphometric uncertainty (Figure 11). First, illumination-induced boundary ambiguity occurred under both low illumination, where the dark body showed limited contrast with the surrounding water, and strong illumination, where body shadows could be confused with the actual contour. Second, under low-surfacing conditions, the apparent boundary could comprise the exposed body, partially visible underwater portions, and water-surface reflections, increasing uncertainty in body-axis estimation and width extraction. Third, splashes and specular reflections could locally obscure or extend the body boundary, resulting in under- or over-segmentation and consequently biased body dimensions. These effects were particularly evident during rapid surfacing and re-entry. Overall, morphometric uncertainty reflected the combined influence of upstream OBB localization, downstream segmentation, imaging conditions, water-surface interference, and transient porpoise posture rather than the performance of a single processing stage.

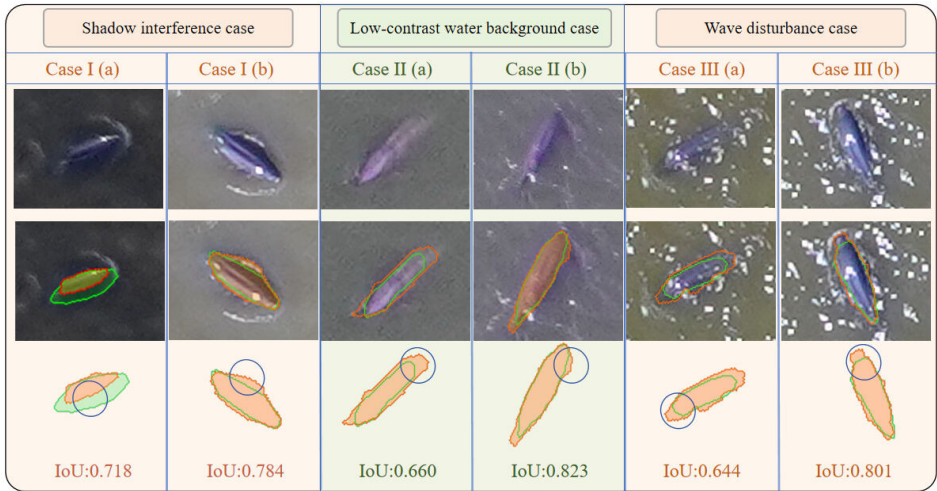


Figure 11. Representative error sources in body-surface morphometric parameter extraction under different UAV imaging and surfacing conditions: Case I, shadow interference; Case II, low-contrast water background; Case III, wave disturbance. For each condition, (a,b) denote two representative samples.

5.3. Implications for UAV-Based Morphometric Monitoring

Continuous-sequence analysis was conducted to evaluate within-event variation and support representative-frame selection. Among 56 surfacing sequences, 8 (14.3%) showed high stability, 23 (41.1%) showed moderate variation, and 25 (44.6%) showed high variation in maximum visible body width (Figure 12). Thus, most sequences exhibited noticeable inter-frame variation because body exposure, posture, and body-axis orientation changed rapidly during surfacing. Treating adjacent frames as independent samples would therefore introduce repeated observations and increase posture-related measurement variability. Accordingly, only one representative frame was retained from each sighting record, with priority given to the image showing the greatest trunk exposure, clearest contour, and most unambiguous body-axis direction. A short image acquisition interval increased the probability of capturing such a measurable frame rather than providing multiple independent morphometric samples from the same surfacing event. The biological plausibility of the extracted measurements was further assessed using previously reported morphometric relationships for finless porpoises. Three-dimensional modeling research has shown that body width and body height can be used to describe body cross-sections, with body width serving as an important input for estimating elliptical cross-sectional girth and body volume [41]. Although maximum visible body width in nadir-view UAV imagery cannot be directly equated with maximum girth or body length, it reflects the transverse scale of the trunk and therefore provides a morphological basis for indirect body-length estimation.

Measurements from seven complete-body samples showed that the ratio of maximum body width to body length ranged from 0.157 to 0.226, with a mean value of 0.197. Given the limited number of complete-body observations, the derived ratio should be regarded as a preliminary indicator for exploring the potential relationship between UAV-derived body width and body length.

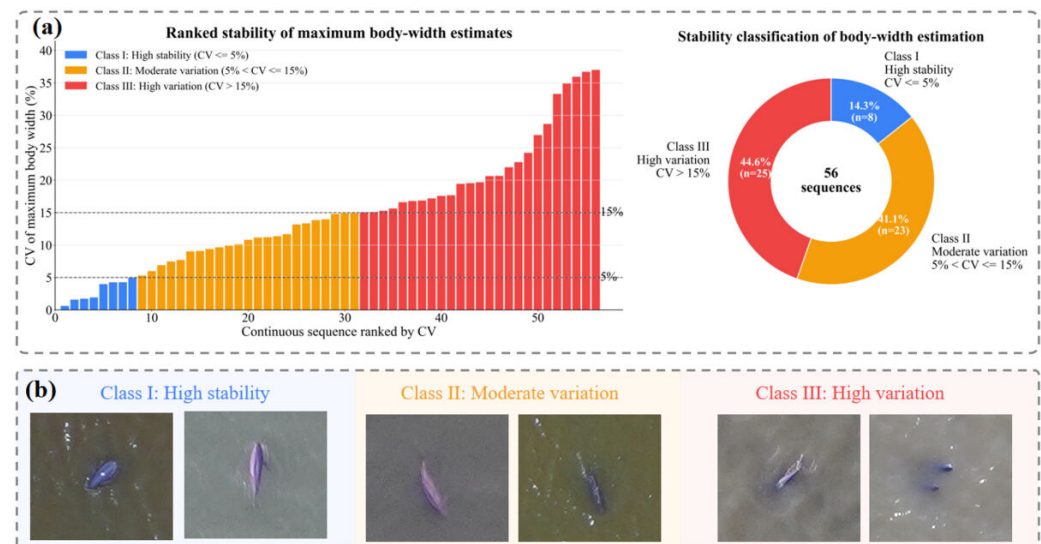


Figure 12. Stability analysis of maximum visible body-width measurements across continuous surfacing sequences: (a) ranked coefficients of variation and stability classification of the 56 continuous surfacing sequences; (b) representative UAV images of Class I (high stability), Class II (moderate variation), and Class III (high variation).

6. Conclusions

A non-contact framework combining YOLO26-OB and SAM2 was developed to extract body-surface morphometric parameters of Yangtze finless porpoises from UAV imagery. YOLO26-OB provided target localization and body-axis information, while

SAM2 generated pixel-level masks suitable for subsequent morphometric measurements. The R-Box + 1FG strategy achieved the best segmentation performance, highlighting the importance of spatial and directional guidance for small targets with weak boundaries. The resulting masks supported the extraction of body length, tail fluke width, maximum body width, and fixed-position body widths from complete-body samples, as well as maximum visible body width from partial-body samples. The automatic measurements were generally consistent with manual measurements. Future work could extend the current image-based framework to UAV videos by incorporating temporal information, individual tracking, and multi-frame contour integration. In addition, the extracted visible body-surface contours and width profiles may provide inputs for future biologically calibrated morphometric and body-condition studies, but should not yet be interpreted as direct indicators of health or reproductive state.

Author Contributions: Conceptualization and methodology, S.W., D.Y. and F.W.; data acquisition, Z.W., T.C. and J.Q.; data curation and validation, X.X. and D.Y.; writing—original draft preparation, S.W., D.Y. and F.W.; writing—review and editing, X.X., S.Z. and F.W.; project administration and funding acquisition, F.W. and S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported by the Central Public-Interest Scientific Institution Basal Research Fund, CAFS (NO. 2024TD04) and the National Non-Profit Institutes (East China Sea Fisheries Research Institute) (2025TD02).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Bierlich, K.; Hewitt, J.; Bird, C.; Schick, R.; Friedlaender, A.; Torres, L.; Dale, J.; Goldbogen, J.; Read, A.; Calambokidis, J. Comparing uncertainty associated with 1-, 2-, and 3D aerial photogrammetry-based body condition measurements of baleen whales. *Front. Mar. Sci.* **2021**, *8*, 749943.
2. Christiansen, F.; Sironi, M.; Moore, M.; Di Martino, M.; Ricciardi, M.; Warick, H.; Irschick, D.; Gutierrez, R.; Uhart, M. Estimating body mass of free-living whales using aerial photogrammetry and 3D volumetrics. *Methods Ecol. Evol.* **2019**, *10*, 2034–2044. <https://doi.org/10.1111/2041-210X.13298>.
3. Qiao, Q.; Zhou, Q.; Wang, J.; Lin, H.; Li, B.; Du, H.; Yan, Z. Environmental DNA reveals the spatiotemporal distribution and migration characteristics of the Yangtze finless porpoise, the sole aquatic mammal in the Yangtze River. *Environ. Res.* **2024**, *263*, 120050.
4. Turvey, S.; Pitman, R.; Taylor, B.; Barlow, J.; Akamatsu, T.; Barrett, L.; Zhao, X.; Reeves, R.; Stewart, B.; Wang, K. First human-caused extinction of a cetacean species? *Biol. Lett.* **2007**, *3*, 537.
5. Mei, Z.; Hao, Y.; Zheng, J.; Wang, Z.; Wang, K.; Wang, D. Population status and conservation outlooks of Yangtze finless porpoise in the Lake Poyang. *J. Lack. Sci.* **2021**, *33*, 1289–1298.
6. Wang, Z.; Tang, B.; Wang, K.; Hao, Y.; Yang, F. Accumulation and risk prioritization of psychoactive substances in the critically endangered Yangtze finless porpoise. *J. Hazard. Mater.* **2023**, *442*, 130002. <https://doi.org/10.1016/j.jhazmat.2022.130002>.
7. Duan, P.X.; Wang, Z.T.; Akamatsu, T.; Tregenza, N.; Li, G.Y.; Wang, K.X.; Wang, D. Anthropogenic activity, hydrological regime, and light level jointly influence temporal patterns in biosonar activity of the Yangtze finless porpoise at the junction of the Yangtze River and Poyang Lake, China. *Zool. Res.* **2023**, *44*, 919.
8. Nabi, G.; Hao, Y.; Zeng, X.; Wang, D. Assessment of Yangtze finless porpoises (*Neophocaena asiaorientalis*) through biochemical and hematological parameters. *Zool. Stud.* **2017**, *56*, e31.
9. Castrillon, J.; Bengtson Nash, S. Evaluating cetacean body condition; a review of traditional approaches and new developments. *Ecol. Evol.* **2020**, *10*, 6144–6162. <https://doi.org/10.1002/ece3.6301>.

10. Hao, Y.J.; Zhao, Q.Z.; Wu, H.P.; Chen, D.Q.; Gong, C.; Li, L.; Wang, D. Physiological responses to capture and handling of free-ranging male Yangtze finless porpoises (*Neophocaena phocaenoides asiaeorientalis*). *Mar. Freshw. Behav. Physiol.* **2009**, *42*, 315–327. <https://doi.org/10.1080/10236240903302161>.
11. Ciciarella, R.; Willems, E.; Markham, B.; Bizzozzero, M.; Phillips, W.; Allen, S.; Krützen, M.; Christiansen, F. Validation of aerial photogrammetry methods to measure body size, condition and mass in small cetaceans. *J. Physiol.* **2026**, *604*, 1788–1808. <https://doi.org/10.1113/JP290419>.
12. Wang, S.; Chen, K.; Wei, Z.; Yang, L.; Wang, Q.; Liu, M.; Cao, K.; Zhao, C.; Chang, R.; Wang, Z. UAV-based deep learning for biodiversity monitoring: Advances, applications, and future directions. *Ecol. Inform.* **2026**, *95*, 103710.
13. Tripathi, R.; Agarwal, K.; Tripathi, V.; Badola, R.; Hussain, S. Conservation in action: Cost-effective UAVs and real-time detection of the globally threatened swamp deer (*Rucervus duvaucelii*). *Ecol. Inform.* **2025**, *85*, 102913.
14. Burnett, J.; Lemos, L.; Barlow, D.; Wing, M.; Chandler, T.; Torres, L. Estimating morphometric attributes of baleen whales with photogrammetry from small UASs: A case study with blue and gray whales. *Mar. Mammal. Sci.* **2019**, *35*, 108–139. <https://doi.org/10.1111/mms.12527>.
15. Dawson, S.; Bowman, M.; Leunissen, E.; Sirguy, P. Inexpensive aerial photogrammetry for studies of whales and large marine animals. *Front. Mar. Sci.* **2017**, *4*, 366.
16. Lu, Y.; Bi, H.; Jiang, W.; Cui, C.; Zhu, J.; Wang, P.; Xu, X.; Yang, G.; Chen, W.; Wang, Z. Two Strategies for Researching the Endangered Yangtze Finless Porpoises Suggest Data-Poor Areas Are Worthy of Greater Conservation Efforts. *Ecol. Evol.* **2025**, *15*, e71649. <https://doi.org/10.1002/ece3.71649>.
17. Corcoran, E.; Anschau, M.; Sudholz, A.; Hamilton, G. Automated detection of wildlife using drones: Synthesis, opportunities and constraints. *Methods Ecol. Evol.* **2021**, *12*, 1103–1114. <https://doi.org/10.1111/2041-210X.13581>.
18. Morimura, N.; Itahara, A.; Brooks, J.; Mori, Y.; Piao, Y.; Hashimoto, H.; Mizumoto, I. A Drone Study of Sociality in the Finless Porpoise *Neophocaena asiaeorientalis* in the Ariake Sound, Japan. *Drones* **2023**, *7*, 422.
19. Xiao, J.; Wang, D. Respiratory pattern of captive Yangtze finless porpoises (*Neophocaena phocaenoides asiaeorientalis*). *J. Ethol.* **2006**, *24*, 205–212. <https://doi.org/10.1007/s10164-005-0181-3>.
20. Delplanque, A.; Linchant, J.; Vincke, X.; Lamprey, R.; Théau, J.; Vermeulen, C.; Foucher, S.; Ouattara, A.; Kouadio, R.; Lejeune, P. Will artificial intelligence revolutionize aerial surveys? A first large-scale semi-automated survey of African wildlife using oblique imagery and deep learning. *Ecol. Inform.* **2024**, *82*, 102679.
21. Xu, Z.; Wang, T.; Skidmore, A.; Lamprey, R. A review of deep learning techniques for detecting animals in aerial and satellite images. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *128*, 103732.
22. Ye, Q.; Ma, M.; Zhao, X.; Duan, B.; Wang, L.; Ma, D. ADD-YOLO: An algorithm for detecting animals in outdoor environments based on unmanned aerial imagery. *Measurement* **2025**, *242*, 116019.
23. Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M. A review of semantic segmentation using deep neural networks. *Int. J. Multimed. Inf. Retr.* **2018**, *7*, 87–93. <https://doi.org/10.1007/s13735-017-0141-z>.
24. Mo, Y.; Wu, Y.; Yang, X.; Liu, F.; Liao, Y. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing* **2022**, *493*, 626–646.
25. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.; Lo, W.Y. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 4015–4026.
26. Ravi, N.; Gabeur, V.; Hu, Y.T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; Rädle, R.; Rolland, C.; Gustafson, L. Sam 2: Segment anything in images and videos. In Proceedings of the International Conference on Learning Representations, Singapore, 24–28 April 2025; pp. 28085–28128.
27. Canelas, J.; Clementino, L.; Cid, A.; Castro, J.; Machado, I.; Vieira, S. Automated cetacean detection in UAV imagery using AI models: A case study on Delphinid species. *Int. J. Data Sci. Anal.* **2025**, *20*, 3965–3979. <https://doi.org/10.1007/s41060-024-00704-9>.
28. Gray, P.; Bierlich, K.; Mantell, S.; Friedlaender, A.; Goldbogen, J.; Johnston, D. Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. *Methods Ecol. Evol.* **2019**, *10*, 1490–1500. <https://doi.org/10.1111/2041-210X.13246>.
29. Torres, W.; Bierlich, K. MorphoMetriX: A photogrammetric measurement GUI for morphometric analysis of megafauna. *J. Open Source Softw.* **2020**, *5*, 1825. <https://doi.org/10.21105/joss.01825>.
30. Bierlich, K.C.; Karki, S.; Bird, C.N.; Fern, A.; Torres, L.G. Automated body length and body condition measurements of whales from drone videos for rapid assessment of population health. *Mar. Mammal Sci.* **2024**, *40*, e13137. <https://doi.org/10.1111/mms.13137>.
31. Bagchi, C.; Medina, J.; Irschick, D.J.; Maji, S.; Christiansen, F. Automated extraction of right whale morphometric data from drone aerial photographs. *Remote Sens. Ecol. Conserv.* **2025**, *11*, 438–453. <https://doi.org/10.1002/rse2.70001>.
32. de Oliveira, L.L.; Andriolo, A.; Cremer, M.J.; Zerbini, A.N. Aerial photogrammetry techniques using drones to estimate morphometric measurements and body condition in South American small cetaceans. *Mar. Mammal Sci.* **2023**, *39*, 811–829. <https://doi.org/10.1111/mms.13011>.

33. Shibata, M.; Ogawa, M.I.; Kimura, S.S.; Yoda, K. The First Estimates of Body Lengths in Free-Ranging Narrow-Ridged Finless Porpoises Using Unmanned Aerial Vehicle Footage. *Mammal Study* **2026**, *51*, 129–136. <https://doi.org/10.3106/ms2025-0023>.
34. Huang, J.; Mei, Z.; Chen, M.; Han, Y.; Zhang, X.; Moore, J.; Zhao, X.; Hao, Y.; Wang, K.; Wang, D. Population survey showing hope for population recovery of the critically endangered Yangtze finless porpoise. *Biol. Conserv.* **2020**, *241*, 108315.
35. Mei, Z.; Zhang, X.; Huang, S.L.; Zhao, X.; Hao, Y.; Zhang, L.; Qian, Z.; Zheng, J.; Wang, K.; Wang, D. The Yangtze finless porpoise: On an accelerating path to extinction? *Biol. Conserv.* **2014**, *172*, 117–123.
36. Roberts, D.; Bahn, V.; Ciuti, S.; Boyce, M.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. <https://doi.org/10.1111/ecog.02881>.
37. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y. Yolo4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
38. Shorten, C.; Khoshgoftaar, T. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. <https://doi.org/10.1186/s40537-019-0197-0>.
39. Jocher, G.; Qiu, J.; Liu, M.; Lyu, S.; Akyon, F.; Kalfaoglu, M. Ultralytics YOLO26: Unified Real-Time End-to-End Vision Models. *arXiv* **2026**, arXiv:2606.03748.
40. Petso, T.; Jamisola, R., Jr.; Mpoeleng, D.; Bennitt, E.; Mmereki, W. Automatic animal identification from drone camera based on point pattern analysis of herd behaviour. *Ecol. Inform.* **2021**, *66*, 101485.
41. Zhang, C.; Zhou, H.; Shah, S.; Davis, R.; Hao, Y.; Yung, K.; Wang, K.; Wang, D. Computer vision-assisted photogrammetry and one-image 3D modeling in marine mammals. *Mar. Mammal Sci.* **2024**, *40*, e13083. <https://doi.org/10.1111/mms.13083>.

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.