Machine Learning (2025) 114:158
https://doi.org/10.1007/s10994-025-06787-0

Towards interpretable and robust UAV-based foundation model for endangered species monitoring in complex ecosystems

Jihan Zhang[1] · Mingqiao Han[1] · K. H. Laurie[2] · Benyun Zhao[1] · Lei Lei[1] · Xi Chen[1] · Hon Chi Judy Wan[3] · Siu Gin Cheung[4] · Wenxing Hong[5] · Ben M. Chen[1]

Received: 1 March 2025 / Revised: 3 April 2025 / Accepted: 23 April 2025

Abstract

Recent advancements in foundation models have significantly enhanced the robustness and scalability of traditional methods in a variety of domains. However, their application to specialized ecological environments, where challenges such as data scarcity, camouflage, and environmental noise persist, remains an area requiring further exploration. This study investigates the application of foundation models in species monitoring within complex ecological systems, with a focus on juvenile Tri-spine horseshoe crabs (*Tachypleus tridentatus*) in Hong Kong's intertidal zones. Traditional methods for monitoring these endangered species are labor-intensive, imprecise, and disruptive to fragile ecosystems, particularly in environments where juveniles exhibit excellent camouflage and small-scale behavioral markers. Unmanned aerial vehicles (UAVs) offer a promising solution, yet their use in these settings is hampered by tidal movements, water turbidity, and complex backgrounds. To address these challenges, we apply a foundation model, Segment Anything Model 2 (SAM2), to UAV-based high-resolution imagery. By leveraging expert knowledge to design and extract domain-specific features, we fine-tune SAM2 using a few-shot learning strategy, enhancing its ability to accurately segment foraging trails with limited data. The fine-tuned model incorporates interpretable morphological features, such as trail length, width, and continuity, to distinguish biological trails from environmental noise, thereby improving both model robustness and interpretability. This approach demonstrates the efficacy of adapting foundation models for domain-specific challenges, advancing both the interpretability and reliability of ecological monitoring systems. The resulting species distribution maps provide valuable insights into population patterns, offering a scalable and transferable solution for monitoring endangered species in dynamic, data-scarce environments. This research highlights the potential of foundation models to revolutionize ecological monitoring by improving model trustworthiness and extending their application to complex, real-world problems.

Editors: Yonggang Zhang, Zhen Fang, Mengyue Yang, Kaifeng Lyu, Hangyu Li, Sharon Li, Kun Zhang.

Extended author information available on the last page of the article

Published online: 05 June 2025

Keywords Foundation model · Expert knowledge · UAV · Aerial surveys · Image processing

1 Introduction

Agricultural production and aquaculture expansion have increasingly transformed intertidal ecosystems to meet the growing global demand for food. In regions such as Hong Kong, where the dynamic interaction between the saline waters of the South China Sea and the freshwater of the Pearl River Delta creates ideal conditions for oyster cultivation, these changes are particularly pronounced (Chan et al., 2022). Oyster farming on the mudflats in Deep Bay in Hong Kong (which includes Ha Pak Nai) has a deep cultural and historical significance and is recognized as part of Hong Kong's intangible cultural heritage (ICHO, 2024). In Hong Kong a traditional method of oyster farming involves bottom planting using rock plates, concrete stakes or large stones deployed in rows as protruding arrays on intertidal flats.

The mudflat at Pak Nai/Ha Pak Nai is of conservation significance, because it hosts the largest population of juvenile Trispine horseshoe crabs (*Tachypleus tridentatus*) in Hong Kong. *T. tridentatus* is a benthic species classified as endangered on the IUCN Red List (Laurie et al., 2019), and it is one of the organisms that is adversely affected by oyster farming. Juvenile *T. tridentatus* depend heavily on sandy intertidal habitats for foraging and protection during their developmental stages. However, habitat loss, fragmentation, and pollution caused by intensive aquaculture practices (including oyster farming) have led to significant population declines. The alteration of sediment structures and tidal regimes further destabilizes these ecosystems (Nordlund et al., 2014).

To address these challenges, habitat restoration initiatives have been launched to rehabilitate degraded mudflats (Fig. 1(a)). Evaluating the success of these efforts requires accurate monitoring of vulnerable species like *T. tridentatus*. Traditional methods, including quadrat sampling and visual searches (Fig. 1(b)), face limitations in detecting camouflaged juvenile individuals, which burrow into sediments and leave only subtle foraging trails (Raffaelli, 1996; Wang et al., 2019). Intertidal environments further complicate monitoring due to extreme abiotic fluctuations (e.g., tidal cycles, sediment shifts) that create microhabitat mosaics (Pennings et al., 2005; Defeo & McLachlan, 2013).

(a) (b)

Fig. 1 **a** Oyster Reefs are under cleaning as the habitat for *T. tridentatus*; **b** The quadrat sampling experiment and the quadrat is highlighted in blue (Color figure online)

Unmanned Aerial Vehicles (UAVs) have emerged as a transformative tool for ecological monitoring in such dynamic environments. By capturing high-resolution georeferenced imagery over large areas, UAVs enable non-invasive, spatially explicit data collection while minimizing ecological disturbance (Dujon & Schofield, 2019; Monteiro et al., 2021). Recent studies demonstrate that UAV-derived datasets can serve as critical inputs for training foundation models in remote sensing and biodiversity monitoring (Zhang et al., 2024; Hong et al., 2024). However, detecting small, camouflaged targets like juvenile *T. tridentatus* remains challenging due to their subtle visual signatures (e.g., faint foraging trails) within complex intertidal substrates (Tuia et al., 2022).

Foundation models pretrained on large-scale multimodal data offer potential solutions. Vision-language models like CLIP have achieved zero-shot recognition of species in camera trap imagery (Fabian et al., 2023), and remote sensing foundation models (e.g., SpectralGPT (Hong et al., 2024)) show promise in habitat mapping. However, three critical gaps hinder their application to UAV-based intertidal monitoring:

- Domain adaptation: Most foundation models are trained on terrestrial or open-ocean datasets, lacking sensitivity to the spectral and spatial patterns of intertidal zones (e.g., tidal signatures, sediment textures) (Li et al., 2024).
- Small-target generalization: Generic object detection architectures struggle to localize cryptic organisms like juvenile *T. tridentatus*, whose visual features occupy limited pixels and blend with background substrates (Mou et al., 2023).
- Data efficiency: Endangered species monitoring often relies on small labeled datasets, which are insufficient for fine-tuning large foundation models without overfitting (Hasegawa & Nakano, 2024).

To address these challenges, we propose a UAV-driven framework that integrates high-resolution aerial imagery with a few-shot learning model specifically designed for juvenile *T. tridentatus* detection. Unlike conventional approaches, our method explicitly encodes ecological prior knowledge (e.g., foraging trail morphology, spatial continuity) into the model architecture, aligning with recent advances in domain-adaptive foundation models (Wang et al., 2025). The framework leverages UAVs' capability to capture fine-grained environmental context (e.g., sediment texture, tidal moisture gradients) that is critical for distinguishing cryptic targets. By combining few-shot learning with interpretable feature engineering, the model achieves robust detection accuracy even with limited annotations, addressing the data scarcity challenge.

This study advances the integration of UAV technology and foundation models in ecological monitoring by: (1) Demonstrating how UAV-derived high-resolution data can enhance foundation models' spatial and spectral awareness in intertidal environments; (2) Proposing a hybrid paradigm that embeds domain-specific ecological knowledge into few-shot learning, bridging the gap between generic pretraining and specialized conservation tasks; (3) Providing a scalable solution for monitoring endangered species in dynamic ecosystems, with implications for biodiversity preservation under anthropogenic pressures.

2 Related works

2.1 *T. tridentatus* conservation

T. tridentatus is an inshore, coastal species with five key stages related to spawning, larval development, juvenile development, sub-adult and adult (Laurie et al., 2019). It is a living fossil with significant ecological and evolutionary value and is now endangered due to habitat degradation, overharvesting, and coastal development (John et al., 2018). Recognized by the IUCN Species Survival Commission (SSC) as a priority for conservation, efforts to protect this species have focused on preserving key habitats such as intertidal spawning and nursery areas (John et al., 2021). These habitats are crucial for reproduction and the early development of juvenile horseshoe crabs, which are vital for maintaining population stability. However, effective conservation measures require reliable population data to understand species distribution and abundance (Wang et al., 2019). In this context, monitoring juvenile *T. tridentatus* in intertidal nursery habitats is particularly important. These individuals are often concentrated in specific areas, making them ideal indicators for assessing overall population health and guiding conservation efforts (Wang et al., 2020). Accurate population assessments can also provide critical insights for coordinating government policies and habitat management strategies.

Based on studies in Hong Kong, including at Ha Pak Nai, juvenile *T. tridentatus* nursery grounds occur on intertidal flats (Chiu & Morton, 2000), where they prefer living on sand dominated mudflats (Kwan et al., 2016). They emerge from the sediment when the substratum is exposed during low tides, where they proceed to forage on the intertidal flat areas covered with a thin layer of surface water, or in pools of standing water in seagrass beds (Zhou & Morton, 2004). The shape of their foraging trails is irregular and distinctive (Chiu & Morton, 2004).

Juvenile *T. tridentatus* exhibit behaviors that make them highly sensitive to habitat characteristics and challenging to monitor. Their tendency to bury themselves in sand for self-protection and their preference for specific sediment types mean that their presence is often concealed (Watanabe et al., 2022). Additionally, surface features such as oyster racks, mudflat patterns, and other environmental factors can interfere with visual detectability (Chan et al., 2022). These behavioral traits may complicate population survey using traditional methods. Volunteer-based monitoring approaches, such as systematic quadrat sampling, belt transects, or random visual searches, are typically employed during spawning seasons but have significant limitations (John et al., 2012; Wang et al., 2019; Wisnewski & Tanacredi, 2022). These methods are time-consuming and prone to errors due to the difficulty of distinguishing juveniles from their surrounding environment. Moreover, the physical act of surveying can disturb the habitat, altering surface features and potentially impacting the behavior of the very species being studied.

The challenges associated with traditional methods underscore the need for more advanced and less invasive monitoring techniques. Juvenile *T. tridentatus*' behavioral and morphological characteristics demand innovative approaches that can address the inherent difficulties of detecting camouflaged species in dynamic intertidal zones. Recent advances in monitoring technologies provide new opportunities to overcome these challenges. Side-scan sonar has been used to locate horseshoe crabs during migration, offering a non-invasive alternative for large-scale surveys (Nagiewicz et al., 2022). However, its application is limited to detecting adult crabs and migration events. Acoustic methods have yet to be effectively adapted for juvenile monitoring in complex intertidal zones. The use of UAVs offers another promising avenue, enabling high-resolution, non-invasive imaging of large intertidal areas while minimizing habitat disturbance (Gonzalez et al., 2016). UAV technology can address many limitations of manual surveys, including their inefficiency and observer bias, while providing georeferenced data for spatially explicit population mapping (Monteiro et al., 2021). In *T. tridentatus* conservation, Koyama et al. have deployed UAV imagery to build the map of habitat, but the recognition work is still conducted manually (Koyama et al., 2020). Automated processing of large volumes of aerial image data for the recognition of juvenile *T. tridentatus* remains an urgent task.

2.2 Vision-based approaches and the emergence of foundation models

Traditional computer vision methods have been pivotal in ecological monitoring tasks such as animal re-identification, habitat mapping, and population assessment (Rees et al., 2018; Schneider et al., 2019; Johnston, 2019; Ravoor & Sudarshan, 2020). Despite notable successes in applying feature-engineering or deep learning techniques (Binder et al., 2012; Unger et al., 2023), many scenarios still present considerable challenges. First, data scarcity frequently arises in endangered-species monitoring, where the collection of extensive, high-quality datasets is impeded by logistical or ecological constraints. Second, environmental complexity (e.g., camouflage, tidal movement, or dense vegetation) makes reliable detection and segmentation more difficult, often exacerbating false positives or negatives (Praveena et al., 2024; Lopez-Marcano et al., 2021). These issues underscore the need for adaptive, robust, and interpretable solutions that can operate effectively with limited domain-specific data.

In recent years, foundation models have shown promise in addressing these limitations by providing large-scale pre-trained representations that can be adapted to a multitude of downstream tasks (Lu et al., 2024; Zhang et al., 2024; Hong et al., 2024; Xiao et al., 2024). Within ecological monitoring, the ability of foundation models to learn generalized semantic features has potential to alleviate data constraints and reduce the need for large, fully labeled training sets (Morera, 2024). For instance, multimodal foundation models have been leveraged for zero-shot recognition of animal species in camera trap imagery (Fabian et al., 2023), while others have incorporated human knowledge to incrementally recognize endangered wildlife with minimal data (Mou et al., 2023). In fisheries management, pre-trained architectures have demonstrated improved robustness in fish recognition tasks compared to conventional supervised methods (Hasegawa & Nakano, 2024), and similar efforts are emerging in smart agriculture for pest detection and plant health monitoring (Li et al., 2024). Beyond image analysis alone, some studies have further integrated textual or spectral modalities (e.g., CLIP-like approaches), broadening the scope for ecological and environmental applications (Wang et al., 2025).

Despite these advances, there remain significant gaps in applying foundation models to highly specialized use cases. Complex, fine-grained features-such as the subtle trails of a camouflaged species-may not be readily captured by broad, generic pre-training (Chen et al., 2024; Kirillov et al., 2023). Moreover, ensuring interpretability is a key concern: large, pre-trained models can behave as opaque black boxes, complicating efforts to validate the correctness of predictions in conservation settings. Achieving robust performance under limited data, while providing transparent reasoning, remains an open challenge (Ravi et al., 2024; Doherty et al., 2024).

In this work, we address the pressing challenge of adapting foundation models for endangered-species monitoring in visually complex, data-scarce ecological settings, with a particular emphasis on the adaptability of our methodology to intertidal environments characterized by diverse and complex conditions. Focusing on the juvenile tri-spine horseshoe crab (T. tridentatus), our approach extends recent research directions on few-shot segmentation and recognition in demanding intertidal habitats. Traditional methods fail to robustly detect subtle, camouflaged trails under varying conditions like algal beds, muddy terrains, and sandy substrates. By building on a high-resolution UAV imaging pipeline, we develop a fine-tuned foundation model Tt-SAM2 through an expert-guided strategy that integrates morphology-informed features specifically designed for these challenging scenarios, thereby achieving robust, interpretable segmentation of subtle foraging trails. We then couple this segmentation with an explainable classification framework that effectively filters environmental noise, ultimately providing accurate distribution maps to support conservation initiatives. Although additional validation in varied ecological settings is essential for broader scalability, our results underscore the exceptional adaptability and reliability of our method specifically within the intricate and dynamic conditions encountered in intertidal monitoring of juvenile horseshoe crabs.

The main contributions of this work are:

- We develop a systematic annotation pipeline for UAV-acquired imagery, underpinned by direct field investigations. This yields a carefully curated, small-scale dataset that captures the juvenile *T. tridentatus*'s subtle behavior and morphology-an essential resource for effective model training under data scarcity.
- Leveraging the annotated dataset, we introduce Tt-SAM2, an explainable few-shot learning approach that highlights biologically meaningful cues. This ensures precise segmentation of juvenile *T. tridentatus* trails-even in environments characterized by camouflage, tidal fluctuations, and other visual complexities.
- We propose an RBF-SVM classifier designed around morphological features, enabling transparent discrimination between genuine trails and environmental artifacts. This step reduces false positives and boosts overall accuracy, preserving the interpretability crucial for ecological decision-making.
- We validate the automated system by cross-referencing UAV-based results with manual field surveys, generating reliable spatial distribution models. These models provide actionable insights into juvenile *T. tridentatus* habitats, informing both targeted conservation strategies and broader ecological research.

By uniting expert-guided feature engineering with a robust foundation model, our framework delivers a replicable and explainable solution for monitoring endangered species in dynamic, real-world ecosystems.

3 Methodology

3.1 Problem description

T. tridentatus is an inshore, coastal species, native to East and Southeast Asia, whose juveniles forage on intertidal mudflats. It has suffered significant population declines due to habitat degradation, pollution, and overharvesting. In Hong Kong, the Ha Pak Nai area (illustrated in Fig. 2) has been proposed as a critical intertidal nursery site, hosting one of the most vital nurseries for juvenile *T. tridentatus* (Lee & Morton, 2016). These mudflats

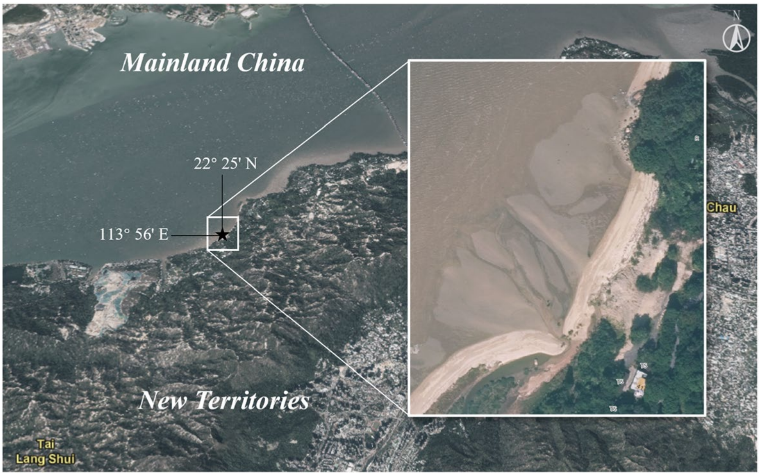


Fig. 2 Horseshoe crab intertidal nursery area, located in Ha Pak Nai

provide the soft sediment and specific ecological conditions-such as tidal rhythms and grain size-necessary for their development. The protection of these habitats is crucial for the survival of this species, as juvenile horseshoe crabs rely heavily on these dynamic environments during early life stages.

For over two decades, the Ocean Park Conservation Foundation has conducted annual surveys in Ha Pak Nai to monitor the population of juvenile *T. tridentatus*. These surveys are performed using quadrat sampling methods (Fig. 1), where volunteers mark specific areas and manually search for horseshoe crabs. While this method has been integral to horseshoe crab research, it suffers from several limitations. First, the rarity and sparse distribution of juvenile horseshoe crabs result in high random and systematic errors, with many surveys yielding few to no observable specimens. Second, comprehensive sampling across the entire mudflat is impractical due to the labor-intensive nature of the method, as well as the temporal constraints imposed by the limited foraging activity periods of juveniles. Lastly, the physical disturbance caused by human presence, including trampling, can alter the surface features of the habitat, disrupt the behavior of juvenile crabs, and hinder conservation efforts.

The challenges of manual sampling are further amplified by the camouflaged behavior and small size of juvenile *T. tridentatus*, as shown in Fig. 3. Juveniles often bury themselves in the sediment as a protective mechanism, leaving only subtle foraging trails as markers of their activity (Fig. 3(b)). These trails, while distinct, are faint and easily obscured by the complex textures of the intertidal zone. The juveniles themselves are exceptionally small, with their body size and trail width often at the centimeter scale (Fig. 3(c)(d)). These factors, combined with the vastness of the habitat, make it exceedingly difficult to accurately and efficiently detect and monitor their populations using traditional methods.

To address these limitations, we propose a UAV-based approach to automate the data collection process and improve monitoring efficiency. As illustrated in Fig. 3(a), UAVs can systematically survey extensive areas of the Ha Pak Nai mudflats, capturing high-resolution aerial imagery without disturbing the habitat. This approach eliminates the need for direct human interaction, thus preserving the natural conditions of the intertidal zone. By leveraging computer vision techniques, including image segmentation and classification,

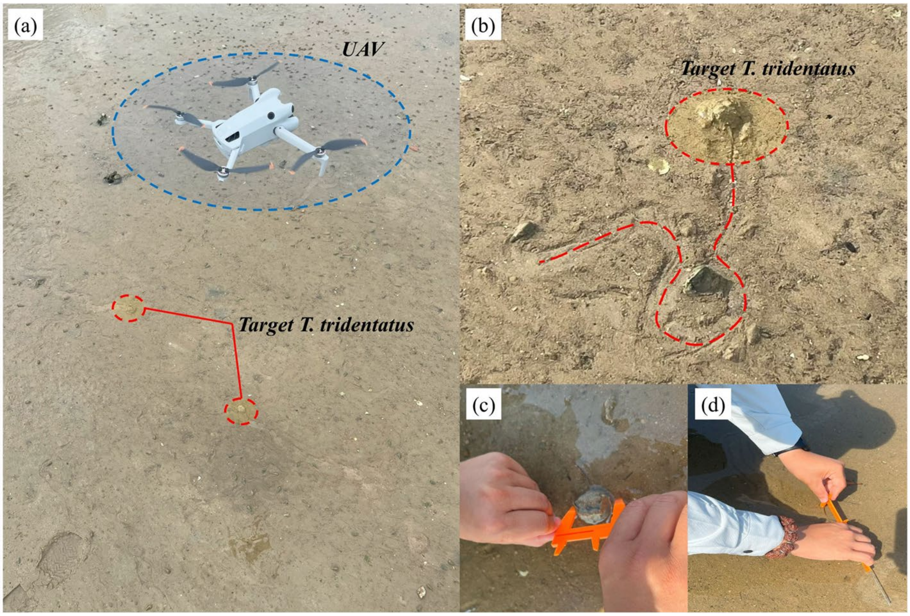


Fig. 3 (a) Real scene of UAV-based data collection on mudflat area (UAV and juvenile *T. tridentatus* are marked with bounding box); (b) The camouflaged juvenile *T. tridentatus* with the typical trail in blue; (c) The measurement of the *T. tridentatus*; (d) The width of the trail at centimeter level (Color figure online)

the UAV imagery is processed to identify juvenile horseshoe crabs and their behavioral markers, such as foraging trails. Our framework integrates expert domain knowledge into the annotation and algorithm design processes, ensuring that the system is capable of identifying subtle features that are often overlooked in traditional methods.

This workflow harnesses UAV-based high-resolution imagery and a foundation-model-driven pipeline to address the limitations of traditional manual surveys (see Fig. 4). Specifically, UAVs systematically collect aerial data over the intertidal habitat, forming the basis for identifying the juvenile *T. tridentatus* and their distinctive foraging trails. To process these data, we employ a fine-tuned foundation model, Tt-SAM2, which was tailored to a small yet highly specialized dataset representing the camouflaged behaviors and morphological nuances of this species. Tt-SAM2 leverages an image encoder, memory attention mechanism, and prompt/mask decoders to segment the potential targets. Subsequently, an explainable classification stage evaluates the segmented trails' morphological attributes to confirm their biological authenticity. This post-segmentation filtering mitigates false positives and culminates in a robust species-distribution map for the intertidal nursery. By automating detection and mapping, our approach not only reduces the labor intensity inherent in manual methods but also offers actionable spatial insights vital for the targeted conservation of *T. tridentatus*.

3.2 Data collection

This study employs UAVs for data collection across the horseshoe crab intertidal nursery area in Ha Pak Nai, minimizing disturbances to the delicate intertidal mudflat

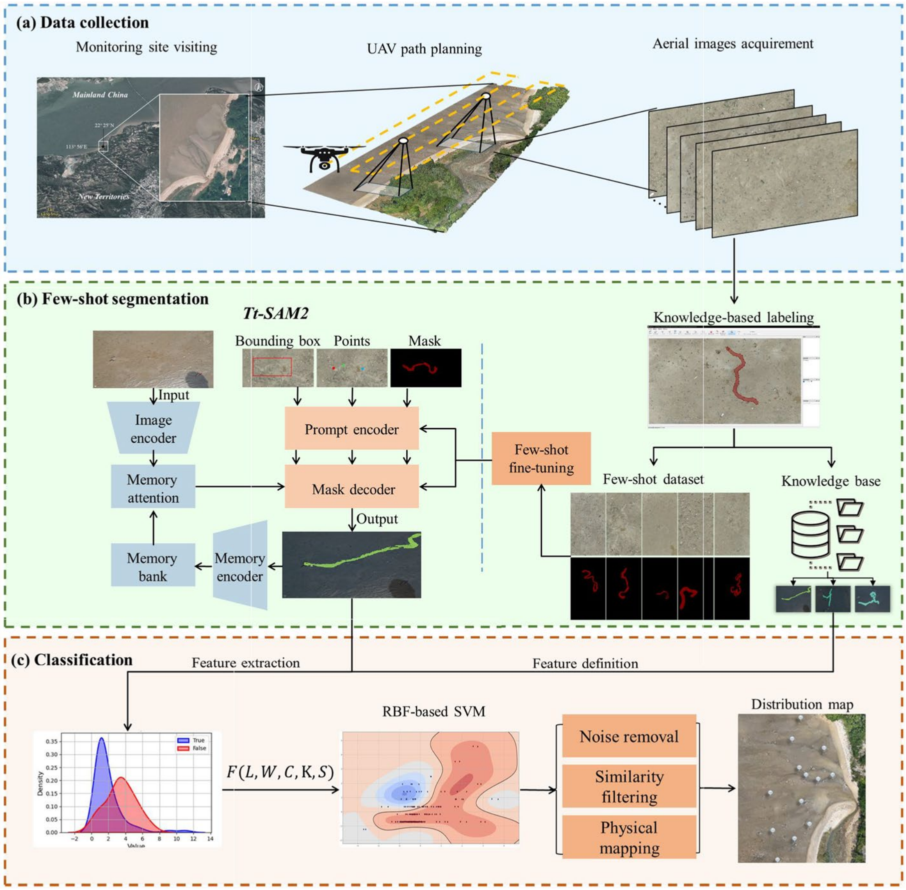


Fig. 4 Overall framework

environment. The use of UAVs eliminates the need for foot traffic, which could disrupt the surface features of the mudflat, potentially harming juvenile *T. tridentatus* and compromising the data quality. UAVs enhance the efficiency and automation of the data collection process, ensuring comprehensive coverage with minimal ecological impact.

To ensure the integrity and completeness of the collected data, the UAV is programmed to follow a carefully designed zig-zag flight path that guarantees full coverage of the designated area while maintaining an overlap rate of $r_o \geq 20\%$ between adjacent flight swaths, as in Fig. 4(a). This overlap ensures that no regions are missed and provides sufficient redundancy for stitching and image analysis. The UAV operates at an average flight altitude of $h = 3\,\mathrm{m}$ above ground level and maintains a constant speed of $v = 5\,\mathrm{m/s}$. To account for the uneven terrain of the mudflat, terrain-following technology is incorporated, ensuring that the UAV maintains a consistent height relative to the surface. Considering the UAV's limited battery life, the total flight distance per mission is planned as $d_f \leq d_{max}$, where d_{max} represents the maximum flight distance achievable within the battery endurance. Upon reaching the battery limit, the UAV is programmed to return to the base for battery replacement before resuming subsequent missions, ensuring continuous and efficient data collection.

The zig-zag flight path is mathematically defined to optimize coverage and ensure data quality. The flight path can be described as a sequence of parallel swaths separated by a distance w, calculated as:

$$w = s \cdot \left(1 - r_o\right), \tag{1}$$

where s is the swath width determined by the UAV's camera field of view, and r_o is the overlap rate. The total flight distance for a single mission, d_f, is given by:

$$d_f = n \cdot \ell, \tag{2}$$

where n is the number of swaths required to cover the target area, and ℓ is the length of each swath. The total number of flights, N, needed to survey the entire area A is computed as:

$$N = \frac{A}{d_{max} \cdot w}. \tag{3}$$

The UAV is deployed with a camera oriented perpendicularly to the ground to capture high-resolution imagery suitable for environmental assessment and juvenile *T. tridentatus* detection. To ensure efficient use of resources, the UAV is programmed to return automatically for battery replacement before resuming the survey. By maintaining a constant altitude and speed, combined with terrain-following capabilities, the UAV collects consistent and high-quality data across the irregular terrain of the mudflat. The proposed path planning method ensures comprehensive coverage of the intertidal zone while minimizing ecological disturbance. The integration of overlap and terrain-following capabilities guarantees data completeness and consistency, even in challenging environmental conditions. This replicable and efficient UAV survey strategy enables periodic monitoring of the horseshoe crab habitat, providing valuable data to support conservation decision-making and long-term ecological studies.

3.3 Few-shot identification with a morphology-centric foundation model

3.3.1 Morphology-guided dataset construction

A key aspect of our methodology involves building a dataset that encodes the distinctive morphological attributes of juvenile *T. tridentatus* foraging trails under data-scarce conditions-a challenge inherent to endangered species monitoring. As shown in Fig. 5, we adopt a multi-phase approach that includes rigorous UAV data filtering, annotation guided by a morphology-oriented knowledge base, and expert verification.

We first preprocess and filter UAV imagery to eliminate low-quality frames (e.g., those blurred by motion or severely occluded). From the filtered images, domain experts identify biologically meaningful patterns, particularly the characteristic three-parallel-line trail structure formed by juvenile *T. tridentatus*. Trained annotators then use LabelMe to precisely delineate these trails, guided by criteria such as spatial continuity, consistent width, and discernible telson traces. Each annotation subsequently undergoes expert validation to

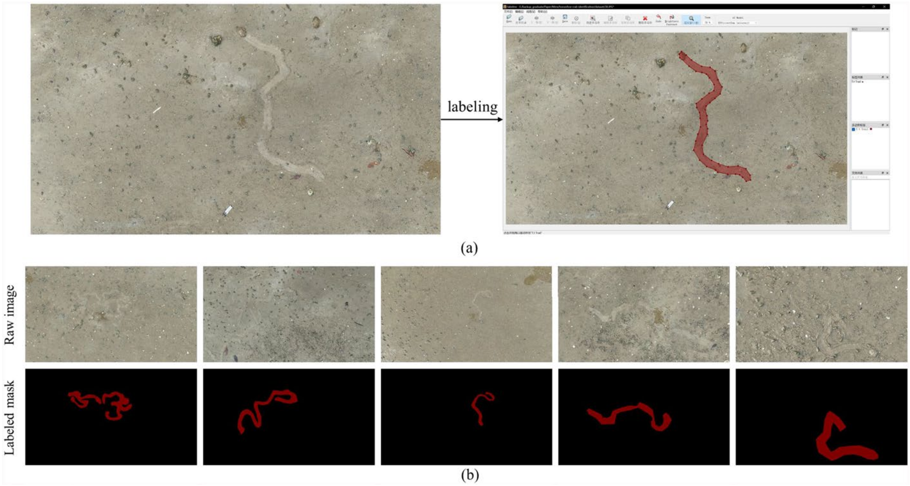


Fig. 5 Illustration of the morphology-based dataset preparation. **a** The multi-phase annotation process, from UAV image filtering to expert validation. **b** Representative samples highlighting the three-parallel-line foraging trails of juvenile *T. tridentatus*

ensure ecological plausibility and to exclude ambiguous samples potentially resulting from environmental noise.

Although the resulting dataset currently consists of only 673 images, each annotation is carefully curated to capture nuanced features such as sediment displacement and fine-scale trail morphology. The dataset's small size is a deliberate reflection of the real-world constraint of limited data availability in endangered species monitoring. Our few-shot learning framework is specifically designed to address this challenge by enabling robust model performance from a limited number of high-quality samples.

Looking forward, we aim to systematically expand this dataset through ongoing UAV-based fieldwork and expert-guided labeling. This planned enrichment will not only support more generalized model training across habitat variability but will also allow us to explore adaptation strategies for scaling our method to other species or intertidal conditions, thereby improving overall robustness and ecological relevance.

3.3.2 Adaptation of a foundation model via few-shot learning

To leverage the representational capacity of large-scale pre-training while accommodating a small, morphology-focused dataset, we adopt the SAM2 (Ravi et al., 2024). Although SAM2 excels in generic segmentation tasks, it requires fine-tuning to capture the nuances of intertidal mudflat imagery and the faint, camouflaged trails of *T. tridentatus*.

Figure 6 outlines the adapted architecture, which retains SAM2's frozen image encoder and optimizes only its prompt encoder and mask decoder. This design minimizes computational overhead and preserves the model's foundational ability to handle diverse visual domains. We incorporate additional morphology-based prompts, such as expected trail width and parallel-line constraints, to guide the segmentation process toward biologically relevant features.

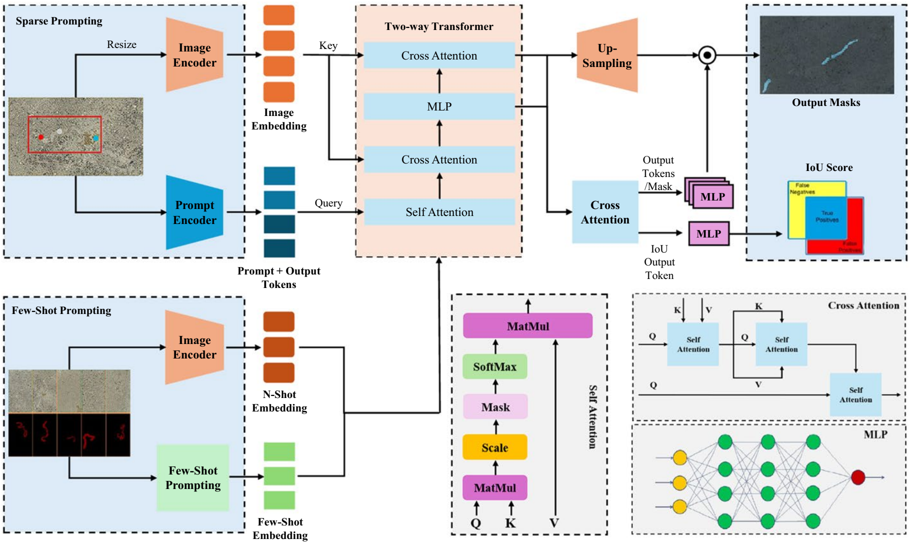


Fig. 6 Overview of the few-shot tuning strategy for the Tt-SAM2 model. The foundation model's image encoder is frozen to preserve general segmentation capabilities, while the prompt encoder and mask decoder are fine-tuned with morphology-based prompts derived from expert knowledge

During fine-tuning, each image is divided into overlapping patches of size $p \times p$ to handle memory constraints associated with high-resolution UAV data. Let $\mathbf{I}_{i,j}$ denote the patch centered at pixel (i, j). Each patch is processed independently, and predictions are later fused to form a global segmentation mask. The training objective is defined by a composite loss function:

$$\mathcal{L} = \lambda_{\mathrm{BCE}}\, \mathcal{L}_{\mathrm{BCE}}(\hat{\mathbf{Y}}, \mathbf{Y}) + \lambda_{\mathrm{dice}}\, \mathcal{L}_{\mathrm{dice}}(\hat{\mathbf{Y}}, \mathbf{Y}) + \lambda_{\mathrm{KL}}\, \mathcal{L}_{\mathrm{KL}}(\hat{\mathbf{Y}}, \mathbf{Y}), \tag{4}$$

where $\hat{\mathbf{Y}}$ and $\mathbf{Y}$ denote the predicted and ground-truth masks for a given patch. The first term, $\mathcal{L}_{\mathrm{BCE}}$, is the binary cross-entropy loss, expressed as

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{N} \sum_{n=1}^{N} \Big[y_n \log(\hat{y}_n) + (1 - y_n) \log(1 - \hat{y}_n) \Big], \tag{5}$$

which penalizes pixel-wise classification errors. The second term, $\mathcal{L}_{\mathrm{dice}}$, measures overlap consistency between predicted and ground-truth segmentation masks, formulated as

$$\mathcal{L}_{\mathrm{dice}} = 1 - \frac{2\,(\hat{\mathbf{Y}} \cap \mathbf{Y})}{\|\hat{\mathbf{Y}}\| + \|\mathbf{Y}\|}, \tag{6}$$

and $\mathcal{L}_{\mathrm{KL}}$ aligns the predicted probability distribution with the ground-truth distribution, mitigating discrepancies in complex scenes.

We optimize the prompt encoder θ_{prompt} and mask decoder θ_{mask}:

$$\theta_{\text{prompt}}, \theta_{\text{mask}} = \arg \min_{\theta_{\text{prompt}}, \theta_{\text{mask}}} \mathbb{E}_{(\mathbf{I},\mathbf{M})\sim\mathcal{D}}\left[\mathcal{L}\big(f(\mathbf{I};\theta_{\text{prompt}}, \theta_{\text{mask}}), \mathbf{M}\big)\right], \tag{7}$$

where $\mathcal{D}$ is the morphology-driven dataset described above, and $f(\mathbf{I};\theta_{\text{prompt}}, \theta_{\text{mask}})$ represents the segmentation output.

Unlike traditional segmentation, where prompts may be generic (e.g., bounding boxes or random points), we incorporate knowledge of *T. tridentatus* morphology. For instance, we provide hints on expected trail continuity and approximate trail width to the model's prompt encoder, thus guiding it to regions more likely to contain target cues. This approach is particularly advantageous for small datasets in which each image is carefully labeled but overall coverage remains limited.

Our methodology contrasts with typical foundation model usage by emphasizing species-specific morphology at every stage. The integration of morphological prompts into a large pre-trained encoder allows Tt-SAM2 to detect subtle foraging trails that might otherwise be overlooked in a generic pipeline. Moreover, patch-based training with overlapping windows ensures local details are captured without sacrificing global contextual cues-crucial for separating authentic trails from environmental textures such as mud ridges or oyster shells.

The resulting Tt-SAM2 model exhibits strong segmentation performance in trials with camouflage or intense background variability, enabling a reliable foundation for downstream classification tasks. This few-shot strategy, grounded in morphology-aware prompts, illustrates how advanced foundation models can be adapted to niche ecological challenges under severe data constraints, providing a scalable template for future conservation-oriented applications.

3.4 Post-segmentation classification

After segmenting potential foraging trails in the UAV imagery, the next crucial step is to confirm whether these segmented trails indeed correspond to juvenile *T. tridentatus* telson traces. This stage combines expert-driven feature engineering with a principled machine learning framework, culminating in an interpretable pipeline that bridges ecological domain insights and automated classification.

A core design principle of our classification framework is the construction of a specialized feature set that embeds domain knowledge about *T. tridentatus* movement and behavior. Specifically, we represent each segmented trail as a five-dimensional feature vector,

$$\mathbf{x} = [L, W, C, \kappa, S],$$

where the five metrics-length (L), width (W), continuity (C), curvature (κ), and shape consistency (S)-together encapsulate the morphological cues that distinguish authentic telson traces from environmental artifacts.

1. Length (L): To measure the overall scale of each foraging trail, we compute the mean length of its contours:

$$L = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \sqrt{\left(x_{j+1} - x_j\right)^2 + \left(y_{j+1} - y_j\right)^2}, \tag{8}$$

where (x_j, y_j) denotes the coordinate of the jth point in the ith contour, M_i is the number of points in that contour, and N is the total number of contours for the trail. We further enhance this measurement by recording the minimum, maximum, and standard deviation of inter-point distances to capture irregularities or abrupt changes. Trails with sufficiently large mean length and small variance in segment length are more likely to correspond to consistent foraging paths.

2. Width (W): To quantify cross-sectional characteristics, we define

$$W = \frac{1}{N} \sum_{i=1}^{N} \frac{A_i}{L_i}, \tag{9}$$

where A_i is the area (in pixels or real-world units) of the ith contour, and L_i is its contour length. This ratio provides a coarse approximation of the trail's transverse dimension. In many cases, actual *T. tridentatus* trails exhibit relatively stable widths, making W and its variance sensitive indicators of authenticity.

3. Continuity (C): Juvenile *T. tridentatus* generally create near-continuous tracks as they traverse the sediment. We compute continuity by counting the number of connected components C_{comp} in the segmented mask. Let $C_{\text{comp}}(\mathbf{M})$ be a function returning the component count of mask $\mathbf{M}$. We then define

$$C = \exp\big(-\alpha\,(C_{\text{comp}}(\mathbf{M}) - 1)\big), \tag{10}$$

where $\alpha > 0$ is a scaling factor. If the segmentation yields exactly one connected component ($C_{\text{comp}} = 1$), then $C \approx 1$. Trails with multiple disconnected blobs produce lower C values, indicating potential environmental noise or overlapping footprints.

4. Curvature (κ): Local bending or turning in the contour can signal interruptions or irregularities. We first compute the first and second derivatives (dx_j, dy_j) and (d^2x_j, d^2y_j) along the contour, then estimate the average curvature as

$$\kappa = \frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \frac{\left| dx_j \cdot d^2y_j - dy_j \cdot d^2x_j \right|}{\left(dx_j^2 + dy_j^2\right)^{3/2}}. \tag{11}$$

We also record the minimum and standard deviation of curvature values, as relatively smooth, consistent arcs align with genuine foraging movements, whereas high curvature variability may indicate noise or sediment disruptions unassociated with crab behavior.

5. Shape consistency (S): To evaluate overall morphological coherence, we examine the difference between each contour's centroid $\mathbf{c}_i$ and its geometric center $\mathbf{g}_i$. Let

$$\mathbf{c}_i = \left(\frac{1}{M_i} \sum_{j=1}^{M_i} x_j,\ \frac{1}{M_i} \sum_{j=1}^{M_i} y_j\right), \quad \mathbf{g}_i = \left(\frac{\max(x) - \min(x)}{2},\ \frac{\max(y) - \min(y)}{2}\right), \tag{12}$$

where $\max(\cdot)$ and $\min(\cdot)$ are computed over the contour coordinates (x_j, y_j). The shape consistency for the entire trail is given by

$$S = 1 - \frac{1}{N} \sum_{i=1}^{N} \mathrm{Var}\big(\mathbf{c}_i - \mathbf{g}_i\big), \tag{13}$$

where a higher S value denotes a contour set whose centroids align well with their geometric centers, indicating a more uniform structure characteristic of a true crab trail.

Before feeding $\mathbf{x} = [L, W, C, \kappa, S]$ into a classifier, we apply domain-specific constraints and statistical normalization. For example, trails with W or L beyond biologically realistic ranges are discarded, and a z-score transform is used to scale each feature to zero mean and unit variance. This procedure mitigates noise effects and ensures consistent weighting across different metrics.

Unlike generic object detection tasks, our morphological feature design is fine-tuned to the idiosyncrasies of a specific endangered species. Each extracted parameter aligns with a documented biological trait, allowing the subsequent classification algorithm to make decisions that are transparent to ecologists. This morphological encoding of domain knowledge stands at the core of our post-segmentation procedure, furnishing an explainable foundation for identifying actual *T. tridentatus* foraging trails with high specificity.

3.4.1 Design of an interpretable classification framework

To transform the extracted feature vectors into final class labels (*true* for juvenile *T. tridentatus* traces vs. *false* for irrelevant patterns), we employ a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel. This choice strikes a balance between interpretability, model capacity, and compatibility with our biology-driven features.

Let $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ denote the dataset of segmented trails, where $\mathbf{x}_i = [L, W, \kappa, C, S]$ is the feature vector for the i-th trail, and $y_i \in \{-1, +1\}$ specifies whether the trail is genuine (+1) or spurious (−1). We learn an optimal decision boundary by solving:

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i, \\ \text{s.t.} \quad & y_i\big(\mathbf{w}\cdot\phi(\mathbf{x}_i) + b\big) \ \geq \ 1 - \xi_i, \quad \xi_i \geq 0, \end{aligned} \tag{14}$$

where $\phi(\mathbf{x}_i)$ is the mapping induced by the RBF kernel:

$$\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) \ = \ \exp\big(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2\big), \tag{15}$$

and γ controls the kernel's degree of nonlinearity. The parameters $\mathbf{w}$, b, and $\boldsymbol{\xi}$ are learned to maximize the margin between the two classes while minimizing misclassifications.

We employ cross-validation to select optimal values for C and γ. Once training is complete, we examine the learned decision function and support vectors to interpret which morphological features most influence trail classification. Large positive weights for length or shape consistency, for instance, reinforce domain expectations that well-defined, continuous structures represent authentic *T. tridentatus* paths. Conversely, a higher reliance on curvature might suggest that sinuous patterns, sometimes caused by water flow or overlapping traces, are more indicative of noise.

Compared to purely deep learning-based pipelines, this morphologically driven classification architecture offers two main advantages:

1. Transparency: By encoding domain knowledge into the feature design and using an SVM with interpretable weights, our model's decisions can be easily traced back to tangible biological reasoning.

2. Robustness under Data Scarcity: The relatively low data requirement for SVM training, combined with carefully curated features, suits rare-species applications where large labeled datasets are unattainable.

Moreover, the feature-engineering approach is adaptable to other camouflaged species or habitats, requiring only a redefinition of morphology-based attributes. Thus, our post-segmentation classification framework not only yields a reliable, domain-consistent interpretation of juvenile *T. tridentatus* trails but also contributes a flexible methodology for broader ecological monitoring contexts.

In summary, the post-segmentation classification stage transforms raw mask outputs from Tt-SAM2 into ecologically valid findings. By synthesizing morphological insight with a margin-based classifier, we ensure high-precision identification of true crab foraging trails. This synergy across domain knowledge, feature design, and an interpretable learning model underpins the robustness and reliability of our overall system, thereby strengthening its potential for real-world conservation applications in dynamic, data-limited environments.

4 Implementation and results

4.1 Experimental setup

The field experiments were conducted in the Ha Pak Nai intertidal nursery area, Hong Kong, during low tide within three hours after the tide receded. This timing coincided with the peak activity of juvenile *T. tridentatus*, ensuring optimal visibility of their trails on the mudflat surface. UAV data collection played a central role in the survey, offering non-invasive data acquisition and comprehensive spatial coverage of the intertidal area without disturbing the habitat. The hardware platform included a UAV equipped with high-resolution imaging sensors and a high-performance workstation for data processing, as summarized in Table 1.

To validate the accuracy and reliability of the UAV-based framework, parallel field experiments were conducted using manual sampling methods. As shown in Fig. 7(b) and (c), experts employed sampling and direct measurements of the trails to verify the presence and dimensions of juvenile *T. tridentatus* activity patterns. These manual surveys provided

Table 1 Specifications of the UAV and workstation used in our experiments

Category	Component	Specification
UAV	Model	DJI Mavic 2 Enterprise Advanced
	Camera Resolution	48 MP RGB Camera
	Flight Altitude	3 ms above ground level
	Flight Speed	5 m/s
	Overlap Rate	20%
Workstation	GPU	NVIDIA A6000
	CPU	Intel® Core™ i9-14900 K
	RAM	128 GB

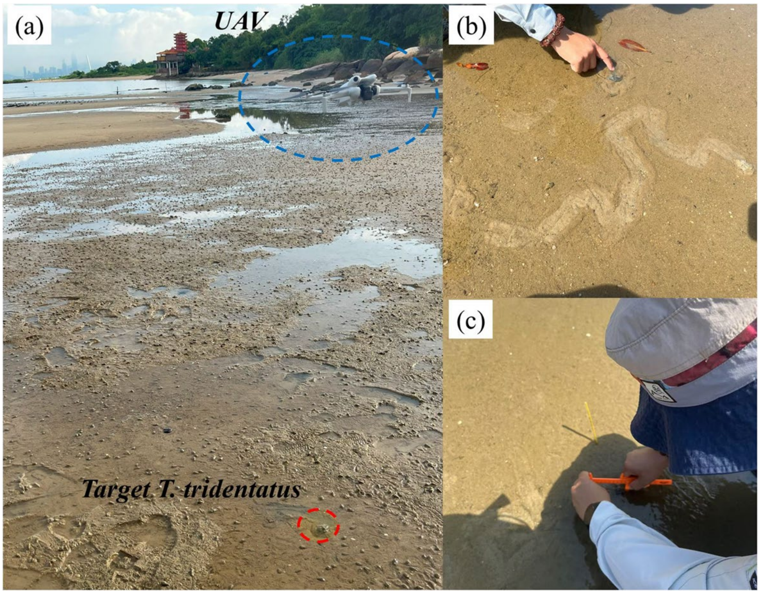


Fig. 7 **a** The UAV data collection in experiment; **b** The parallel study by expert manual method; **c** The measurement of the trails for validation

ground-truth data, enabling cross-verification of the UAV-derived results and ensuring the robustness of the automated detection system.

In terms of efficiency, we conducted a quantitative comparison between the UAV-based and traditional visual search approaches. The manual method relied on human visual search conducted by a team of three experts and several trained volunteers, covering the same survey area over 3.5 h. In contrast, the UAV system-operating autonomously at a flight speed of 5 m/s and an altitude of 3 m-required only 50 min to capture the full coverage of the designated area. The subsequent image segmentation and classification process took an additional 2 min, resulting in a total survey time of approximately 52 min. This reflects a 75.24% reduction in operational time compared to the manual process.

Furthermore, the UAV approach offers substantial cost and labour savings. Manual surveys depend heavily on skilled personnel and coordinated team logistics, whereas UAV flights can be executed autonomously with minimal supervision. This operational advantage is particularly beneficial in remote or expansive coastal zones where repeated field access is time-consuming and labour-intensive.

The integration of UAV-based and manual methods allowed for a comparative analysis, demonstrating not only the alignment of automated detection with traditional ecological survey standards, but also the clear practical benefits of adopting UAV technology in terms of both efficiency and scalability.

4.2 Segmentation results

Our fine-tuned Tt-SAM2 model was evaluated against several state-of-the-art (SOTA) segmentation approaches, including DINO-v2 (Oquab et al., 2023), SOLO-v2 (Wang et al.,

Table 2 Performance metrics for segmentation models

Model	Pixel Accuracy	Dice Coefficient	IoU	Precision	Recall	F1-Score	Result
DINO-v2	–	–	–	–	–	–	Failed
SOLO-v2	–	–	–	–	–	–	Failed
SAM-Auto	–	–	–	–	–	–	Failed
SAM-Prompt	–	–	–	0.0523	–	–	Failed
SAM2-Auto	–	–	–	–	–	–	Failed
SAM2-Prompt	–	–	–	0.1082	–	–	Failed
Ours	0.9621	0.9456	0.8128	0.9969	0.9621	0.9620	Successful

2020), and different variants of SAM (Kirillov et al., 2023). As summarized in Table 2, none of these competing methods produced sufficiently accurate masks for the subtle *T. tridentatus* foraging trails. And the visual comparisons are as in Fig. 8 to illustrate the segmentation results. In contrast, our Tt-SAM2 achieved consistently high scores on key metrics such as Pixel Accuracy (0.9621) and Dice Coefficient (0.9456), underscoring its robustness and adaptability in the challenging mudflat context.

We also tested YOLO-based detection models (e.g., YOLOv10 (Wang et al., 2024)), which reached a maximum mAP_{50} of only 0.0533, highlighting the inherent difficulty of detecting such faint, elongated features. These poor performances strongly indicate that traditional segmentation or detection frameworks struggle in data-scarce, visually noisy environments without explicit ecological or morphological insights. Our Tt-SAM2 solution overcomes these barriers by leveraging domain-specific knowledge, such as the signature three-line patterns and consistent trail continuity of juvenile horseshoe crabs, and embedding these clues into the fine-tuning pipeline.

Figure 9 illustrates how Tt-SAM2 handles various foraging-trail configurations, including curved, crossing, and hybrid patterns. Notably, the ability to capture these fine-grained structures stands as both a technical and ecological advancement: it reveals subtle behavioral differences under varying mudflat surface conditions, potentially leading to new ecological insights on *T. tridentatus* locomotion. These findings reinforce the utility of combining a powerful foundational segmentation model with well-curated, domain-driven fine-tuning strategies, particularly in small-scale ecological monitoring scenarios.

4.3 Post-segmentation classification results

After obtaining the segmented masks from Tt-SAM2, we further classified each trail to confirm whether it represented a valid juvenile *T. tridentatus* foraging pattern. This two-stage approach-segmentation followed by classification-aims to reduce false positives, enhance interpretability, and provide high-resolution distribution maps. Figure 10 presents the correlation matrix for the five morphological features (length L, width W, continuity C, curvature κ, and shape consistency S). Overall, C and S exhibit minimal correlation, signifying their complementary roles in capturing different facets of crab foraging structures.

Figure 11 shows the distributions of L, W, κ, C, and S for both true and false trails. True trails tend to exhibit relatively large L and near-unity C, which aligns with domain knowledge indicating that juvenile crabs often leave continuous and comparatively long paths in

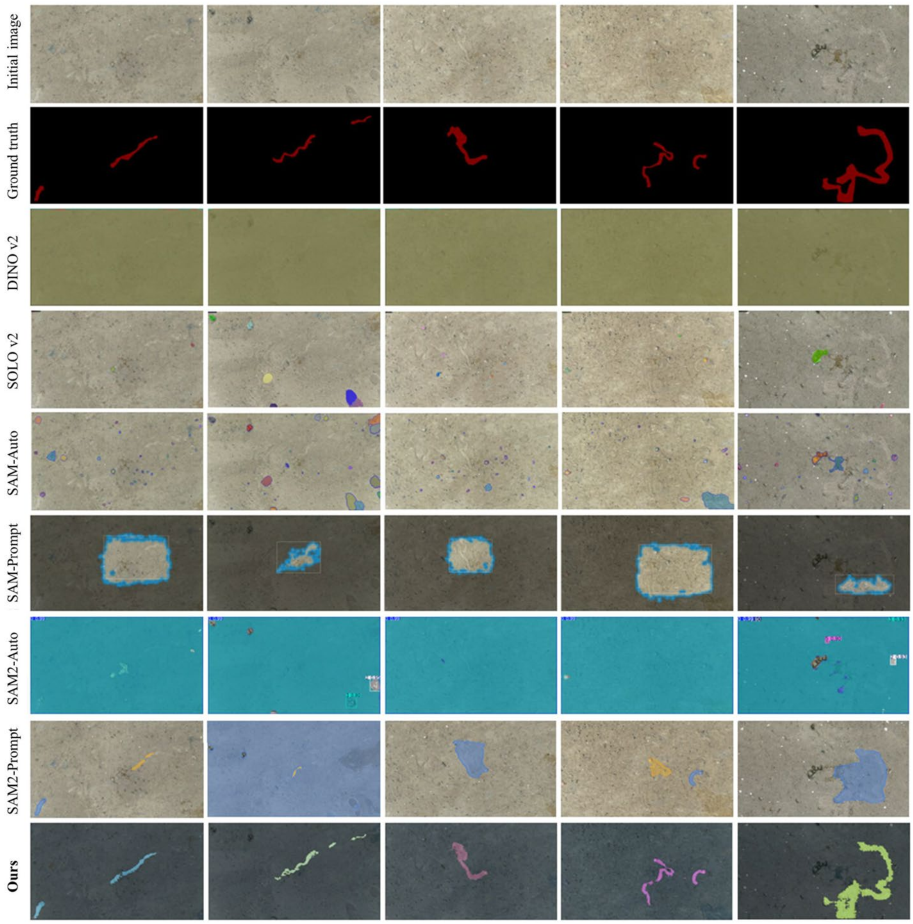


Fig. 8 Visual comparisons with existing segmentation methods. Models like DINO-v2 and SOLO-v2 fail to capture fine-grained foraging trails, while our Tt-SAM2 accurately segments the subtle telson traces

the sediment. Conversely, false trails reveal more irregular patterns in these features, often due to overlapping footprints or environmental noise like debris.

To evaluate the utility of each feature subset, we trained and tested our SVM-based classifier on progressive combinations of L, W, C, κ, and S. As seen in Fig. 12, the full feature set offers the best performance, reaching a maximum accuracy of 93%. This outcome validates our design decision to incorporate multiple domain-informed metrics rather than relying on a single morphological cue.

In addition, we conducted a systematic comparison between our RBF-kernel SVM classifier and several commonly used alternatives, including linear SVM, random forest, k-nearest neighbors (k-NN), logistic regression, and decision tree models. As shown in Table 3, RBF-SVM outperformed all competing classifiers, achieving the highest mean accuracy (94.74%) and F1-score (96.50%) under a 5-fold cross-validation scheme. The superior performance is attributed to the nonlinear mapping capacity of the RBF kernel,

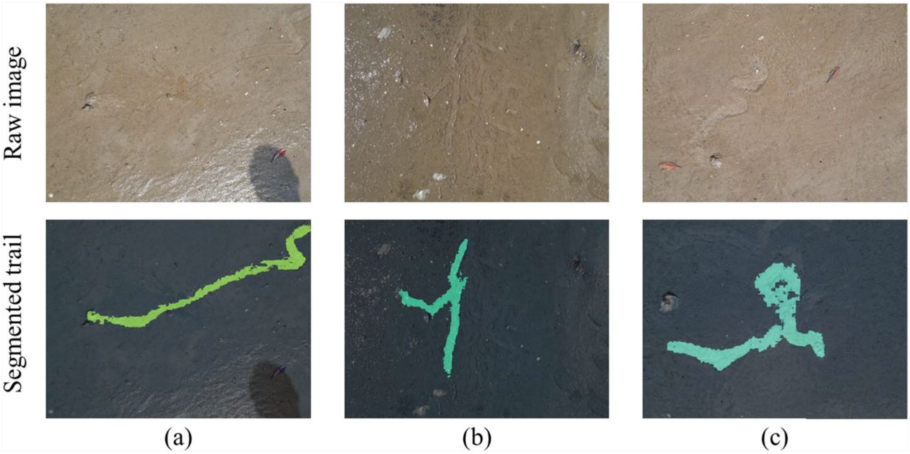


Fig. 9 Exemplary results from Tt-SAM2 segmentation, showing three different trail morphologies: **a** curved trails, **b** crossed trails, and **c** combined morphologies

Fig. 10 Correlation matrix of the extracted features, illustrating pairwise relationships and highlighting potential collinearities

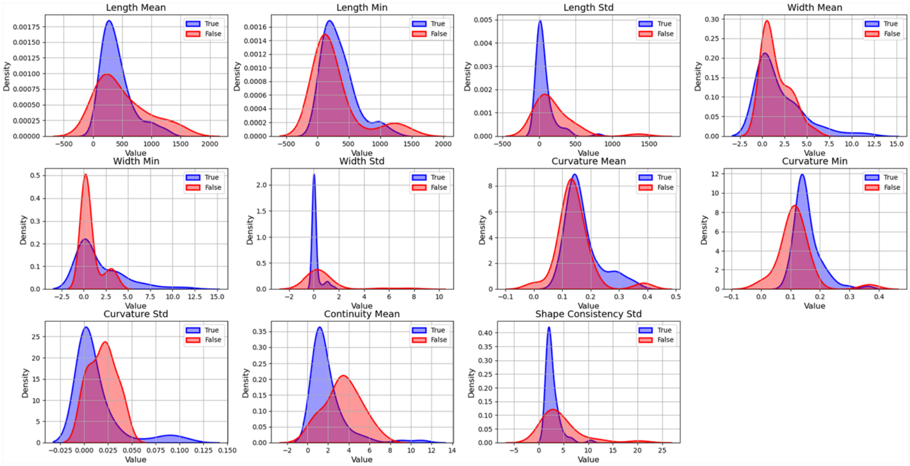


Fig. 11 Feature distributions for true and false trails, with marked separations in L, C, and S. These distinctions highlight the importance of domain-informed feature engineering

which enables more flexible decision boundaries for complex trail morphologies and subtle feature interactions. In contrast, linear models or shallower tree-based classifiers showed reduced sensitivity to the domain-informed features, particularly in ambiguous or noisy regions. These findings further justify our choice of RBF-SVM as the optimal classifier in our pipeline.

To address the impact of hyperparameter tuning in RBF-SVM, we performed a comprehensive grid search over a wide range of values: $C \in \{1, 10, 100\}$ and $\gamma \in \{0.001, 0.01, 0.1, 1, \texttt{scale}\}$, using 5-fold cross-validation. This allowed us to explore the effects of both regularization and kernel smoothness on classification performance. The results, visualized in Fig. 14, show that the optimal hyperparameter configuration

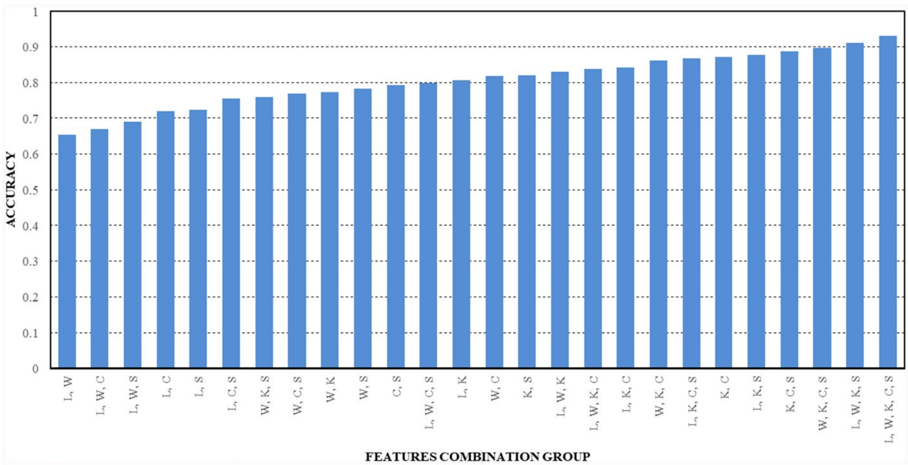


Fig. 12 Classification accuracy achieved by various feature subsets. The complete set (L, W, C, κ, S) yields the highest accuracy of 93%

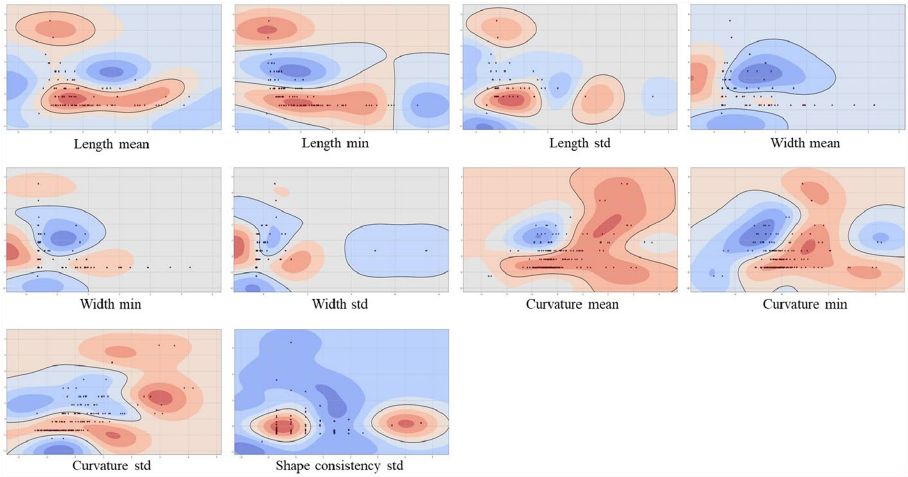


Fig. 13 Decision boundary visualized for the RBF SVM in a reduced feature space. True trails (red) and false trails (blue) show clear separability, underscoring the strength of our morphology-based features (Color figure online)

Table 3 Classification performance comparison using 5-fold cross-validation

Classifier	Mean Accuracy	Mean F1-score
SVM (RBF)	0.9474	0.9650
SVM (Linear)	0.8842	0.9253
Random Forest	0.8737	0.9202
k-NN	0.7684	0.8449
Logistic Regression	0.8632	0.9090
Decision Tree	0.8105	0.8718

was $C = 100$, $\gamma = 0.01$, under which the highest cross-validation accuracy was achieved. We therefore adopted this setting for all subsequent experiments.

Figure 13 illustrates the resulting decision boundary for the RBF-SVM in a two-dimensional projection. Notably, there is a well-defined separation between true and false trails, which highlights the effectiveness of our morphology-focused feature engineering. The smooth gradients at the boundary suggest that the classifier handles instances of partial ambiguity-where trails share attributes of both classes-without abrupt misclassifications.

Finally, we incorporate a post-classification filtering stage to consolidate spatially and morphologically redundant trails, as shown in Fig. 15. This step helps maintain clean distribution maps, preventing overestimation of the *T. tridentatus* population and mitigating noise introduced by multiple UAV passes over the same region. While failure cases can arise under extreme turbidity, highly cluttered substrates, or in the presence of visually similar natural structures, our two-stage design improves generalizability across diverse intertidal environments. The segmentation stage localizes a broad set of candidate trails, including those partially obscured or less distinct, and the classification stage

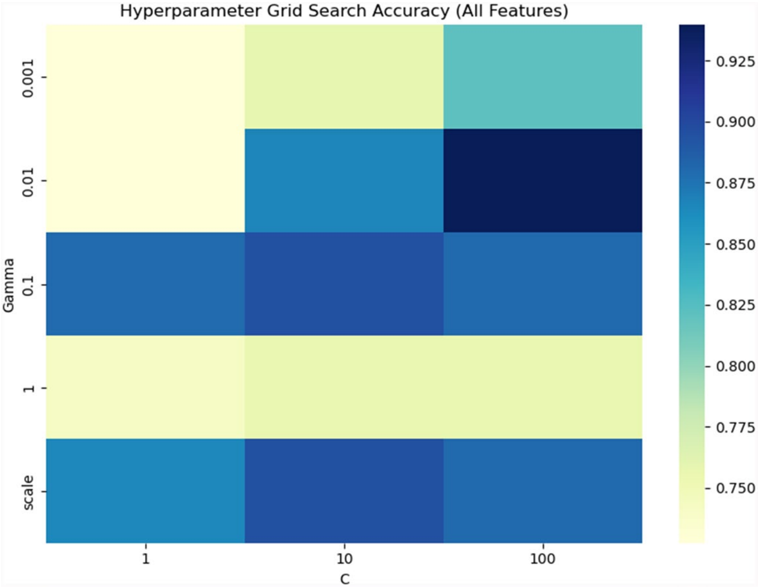
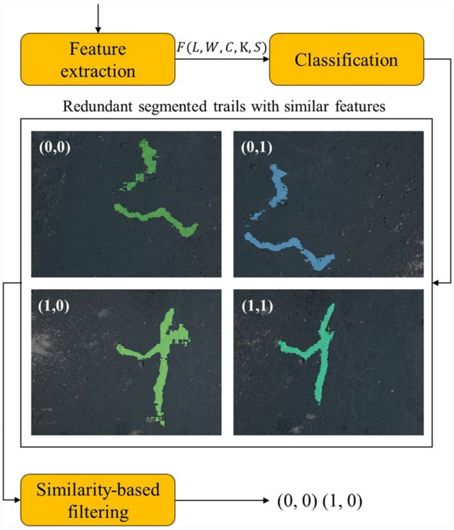


Fig. 14 Heatmap of cross-validation accuracy under different SVM hyperparameters (RBF kernel). Darker regions indicate higher accuracy. The best performance was achieved at $C = 100$, $\gamma = 0.01$

Fig. 15 Feature similarity filtering applied post-classification to remove redundant or overlapping segments in final distribution maps



effectively filters out most non-biological artifacts by enforcing morphological consistency. Empirically, we observe that even in challenging scenarios where features like L and C are less prominent, the classifier maintains good accuracy and avoids systematic misclassifications. Overall, these classification experiments underscore both the accuracy and interpretability of our pipeline, which systematically weaves domain expertise into advanced segmentation and classification components. The end result is a robust,

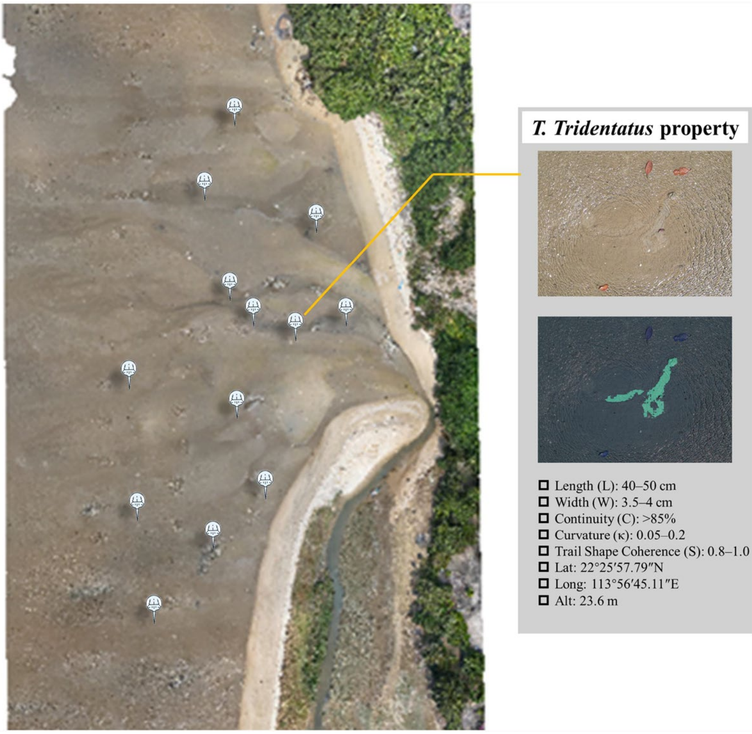


Fig. 16 *T. tridentatus* overlaid on the reconstructed intertidal zone model. Each marker represents an individual trail, providing access to corresponding ecological data such as segmentation results and physical features

ecologically aligned framework capable of identifying juvenile horseshoe crab foraging trails with high precision, even under constraints imposed by limited datasets and challenging intertidal environments.

4.4 Spatiotemporal mapping and ecological data integration

To analyze and visualize the spatial distribution of *T. tridentatus* trails, we developed a comprehensive workflow that integrates UAV-acquired images, geospatial data, and segmentation results. The workflow enables precise identification, mapping, and management of ecological data for conservation purposes. The final results, shown in Fig. 16, present the spatial distribution of juvenile *T. tridentatus* and provide access to detailed ecological records.

The proposed workflow extracts geospatial metadata, maps image coordinates to physical dimensions, reconstructs a 3D scene model, and annotates it with the identified *T. tridentatus* trails. The following pseudocode describes the main steps of this process:

Algorithm 1 Spatiotemporal mapping and data integration workflow

Input: UAV imagery dataset $\mathcal{D}$, camera parameters $\mathcal{C}$, field of view (FOV), flight altitude h
Output: Updated 3D model with annotated *T. tridentatus* trails and synchronized ecological knowledge base
Initialize an empty database $\mathcal{DB}$
Connect to the centralized ecological knowledge base $\mathcal{KB}$
for all image I in $\mathcal{D}$ **do**
    Extract GPS metadata: latitude (ϕ), longitude (λ), altitude (h)
    Perform segmentation on I to extract trail features: L, W, C, κ, S
    $\text{PhysicalScale} \leftarrow \frac{\text{FOV}}{\text{ImageResolution}} \times h$
    Map trail features to real-world dimensions using PhysicalScale
    $\mathcal{DB} \leftarrow \mathcal{DB} \cup \{\phi, \lambda, h, L, W, C, \kappa, S, I\}$
end for
Reconstruct a 3D scene model from the dataset $\mathcal{D}$
Overlay the trail information from $\mathcal{DB}$ onto the georeferenced 3D model
Synchronize the annotated 3D model and trail data to the knowledge base $\mathcal{KB}$

Each element of the pseudocode is designed to translate field-collected UAV data into a spatially indexed and semantically annotated ecological model. The UAV image dataset $\mathcal{D}$ consists of geotagged images that are first associated with metadata retrieved from the image EXIF, notably geographic coordinates (ϕ, λ) and flight altitude h. This metadata enables spatial registration of extracted features.

The segmentation model is then applied to each image $I \in \mathcal{D}$ to extract a set of morphological descriptors: trail length L, width W, continuity C, curvature κ, and shape consistency S. These features reflect ecologically meaningful aspects of *T. tridentatus* movement patterns. To convert image-based features into real-world dimensions, we compute a physical scale factor:

$$\text{PhysicalScale} = \frac{\text{FOV}}{\text{ImageResolution}} \times h, \tag{16}$$

which accounts for the camera's field of view and UAV flight altitude. This scaling allows pixel-based measurements to be projected onto geographic space.

The transformed records are organized in a structured database $\mathcal{DB}$ containing geospatial and morphological attributes linked to each UAV frame. These entries serve as input to the final reconstruction stage, where UAV images are used to generate a 3D terrain model of the study site. This model can be reconstructed using standard photogrammetry software (e.g., Meshroom or Agisoft Metashape) or substituted with a georeferenced 2D map in practical scenarios where full 3D rendering is not required.

The database entries $\phi, \lambda, h, L, W, C, \kappa, S, I$ are subsequently overlaid on the reconstructed model to provide spatial context to trail annotations. Finally, the result is synchronized with a centralized ecological knowledge base $\mathcal{KB}$, supporting long-term storage, retrieval, and integration with other conservation datasets. This modular architecture ensures interpretability and scalability while facilitating the ecological interpretation of spatial patterns in horseshoe crab behavior.

Figure 16 illustrates the spatial distribution of juvenile *T. tridentatus* overlaid on the reconstructed terrain model. This visualization provides critical support for conservation efforts by enabling comprehensive spatial analysis of *T. tridentatus* populations; access to individual trail data, including images, segmentation results, and feature metrics; and integration with the knowledge base to inform adaptive conservation strategies. To strengthen the validity of the study, parallel verification was conducted through manual field surveys in addition to UAV-based data collection. Expert-guided ground-truthing, using traditional methods such as direct visual inspections, was employed to cross-validate the segmentation and classification results obtained from UAV imagery. These parallel approaches have confirmed the accuracy of the UAV-based system in identifying juvenile *T. tridentatus* and distinguishing relevant features from environmental noise, ensuring that the results are robust and reliable. Additionally, the workflow accounts for redundancies caused by UAV data overlap. Redundant trails, representing the same individual, are identified and filtered based on geographic proximity and feature similarity, ensuring accurate population estimates. The processed data is synchronized with a centralized ecological knowledge base, allowing seamless access to all recorded *T. tridentatus* information. The knowledge base supports long-term monitoring, adaptive habitat management, and decision-making for intertidal conservation.

By incorporating spatial, physical, and ecological dimensions, this integrated approach provides a robust foundation for studying and protecting juvenile *T. tridentatus*. The combination of advanced UAV-based imaging, segmentation, and knowledge base updates ensures scalable and replicable conservation strategies for endangered intertidal species.

5 Conclusion

This study presents a novel and comprehensive framework for monitoring the endangered juvenile *T. tridentatus* in visually complex, data-scarce intertidal environments. By uniting UAV-based data collection, a few-shot fine-tuning of the Tt-SAM2 foundation model, and an expert-informed SVM classification stage, we bridge recent advances in machine learning with the pressing needs of ecological conservation. In doing so, our approach not only surpasses the limitations of traditional, labor-intensive surveys but also establishes a meaningful benchmark for leveraging foundation models in specialized ecological contexts.

A key contribution lies in our successful adaptation of the Tt-SAM2 model-a general-purpose foundation model-for the subtle and nuanced task of segmenting *T. tridentatus* trails. Through targeted few-shot learning, guided by morphological and behavioral insights, we achieved robust segmentation performance despite a relatively small dataset of 673 high-resolution images. This underscores the adaptability of foundation models to niche domains when domain-specific knowledge is systematically integrated. Furthermore, the subsequent classification step, driven by an SVM that encodes biologically relevant features (e.g., length, width, continuity), ensures that genuine foraging trails are clearly distinguished from noise. This synergy of segmentation and post-processing exemplifies how computational methods can be harmonized with expert ecological knowledge to produce interpretable and accurate outcomes.

An additional innovation is the integration of segmentation outputs with a 3D ecological knowledge base, enabling spatial visualization and broader contextual analysis of *T. tridentatus* habitat use. This ecosystem-level perspective not only aids real-time monitoring but also fosters iterative knowledge refinement, wherein feedback from ecological experts can further calibrate both the model and the underlying dataset. Such a bidirectional flow of information demonstrates the transformative potential of combining cutting-edge computational frameworks with scientific field expertise, ultimately accelerating the discovery and preservation of threatened species.

Despite its demonstrated strengths, our framework still faces certain challenges that present avenues for future work. Expanding the dataset to encompass multiple habitats and broader environmental gradients would enhance the generalizability of the foundation model, particularly in light of natural variations such as tidal fluctuations and sediment composition. Incorporating temporal monitoring components could also deepen insights into the long-term population dynamics of juvenile *T.tridentatus*, informing more adaptive conservation measures. We thank the reviewer for highlighting the importance of time-series variation. As noted in Sect. 4.1, the current data collection is focused on low tide windows within three hours after tide recession, when juvenile *T. tridentatus* are known to exhibit peak foraging activity. This targeted strategy maximizes observed individuals during each field session. Nevertheless, we acknowledge that juvenile behavior and distribution may vary across tidal cycles, seasons, and environmental rhythms. In future work, we plan to integrate temporal variation into our current spatially structured framework by designing systematic, long-term UAV monitoring schedules. This will enable us to capture behavioral dynamics over different timescales, leading to more comprehensive ecological insights and enhancing the utility of our system for conservation planning. Additionally, exploring multi-modal data sources-ranging from thermal and hyperspectral imaging to acoustic sensors-could further bolster model robustness against the diverse ecological factors that shape intertidal zones.

Looking ahead, our framework holds promise for broader application beyond *T. tridentatus*. Similar morphology-driven fine-tuning procedures can be applied to other endangered species with cryptic behaviors or camouflaged appearances, highlighting the flexibility and scalability of foundation models when reinforced by domain knowledge. In particular, any species requiring large-scale surveys-such as nesting sea turtles, migratory birds, or invasive coastal invertebrates-can benefit from the same pipeline. By incorporating species-specific ecological traits into the annotation phase (e.g., nest morphology, movement traces, or spatial clustering), the segmentation-classification workflow can be adapted to automatically detect and monitor target populations. In many cases, species with more distinct ecological markers may enable even higher automation performance than *T. tridentatus*. This highlights the potential of our approach as a generalized framework for biodiversity surveys, especially in challenging environments where manual data collection is costly or infeasible. Moreover, user-friendly interfaces for data processing and annotation would lower the barrier to adoption, empowering local communities and conservation practitioners to operationalize these computational tools directly in the field.

In summary, this work advances both the methodological frontier of foundation model adaptation and the practical imperative of endangered-species monitoring. By interweaving UAV technology, explainable machine learning, and ecological expertise, we offer a sustainable and high-precision solution for tracking cryptic species in harsh, data-limited habitats. Our contributions pave the way for further integration of foundation models into ecological studies, encouraging a future in which AI-driven insights and biological knowledge coalesce to protect biodiversity and deepen our understanding of complex natural systems.

Author contributions Jihan Zhang and Mingqiao Han led the overall study conception, investigation, and formal analysis. K. H. Laurie, Benyun Zhao and Lei Lei supported conceptualization and data validation. Xi

Chen supervised the project and provided resources, while Hon Chi Judy Wan, Siu Gin Cheung, and Wenxing Hong contributed to conceptualization and manuscript revisions. Ben M. Chen secured funding, offered resources, and oversaw the research direction. All authors reviewed and approved the final manuscript.

Funding This work was supported in part by the InnoHK of the Government of the Hong Kong Special Administrative Region via the Hong Kong Centre for Logistics Robotics (HKCLR), in part by the Environment and Conservation Fund, Hong Kong SAR (Project No. 142/2023), and in part by the Research Grants Council of Hong Kong SAR under Grants 14206821, 14217922, and 14209623.

Data availability Data will be made available upon reasonable request.

Code availability The code of this work will be available after publication.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval This research contains no elements requiring ethical approval.

Consent for publication This research contains no elements requiring consensus.

References

Binder, A., Müller, K.-R., & Kawanabe, M. (2012). On taxonomies for multi-class image categorization. *International Journal of Computer Vision, 99*, 281–301.

Chan, H., Lo, H., Ng, S., Chen, R.-F., & Cheung, S. (2022). Visually guided behaviour of juvenile tri-spine horseshoe crab tachypleus tridentatus: Implications for designing racks for oyster cultivation compatible with horseshoe crab conservation. *International Horseshoe Crab Conservation and Research Efforts: 2007–2020: Conservation of Horseshoe Crabs Species Globally* (pp. 177–187). Springer.

Chiu, H. M., & Morton, B. (2000). The distribution of horseshoe crabs (tachypleus tridentatus and carcinoscorpius rotundicauda) in Hong Kong. *Asian Marine Biology, 16*, 185–196.

Chiu, H. M., & Morton, B. (2004). The behaviour of juvenile horseshoe crabs, tachypleus tridentatus (xiphosura), on a nursery beach at shui hau wan, hong kong. *Hydrobiologia, 523*, 29–35.

Chan, S. S., Wong, H. T., Thomas, M., Alleway, H. K., Hancock, B., & Russell, B. D. (2022). Increased biodiversity associated with abandoned benthic oyster farms highlight ecosystem benefits of both oyster reefs and traditional aquaculture. *Frontiers in Marine Science, 9*, Article 862548.

Defeo, O., & McLachlan, A. (2013). Global patterns in sandy beach macrofauna: Species richness, abundance, biomass and body size. *Geomorphology, 199*, 106–114.

Dujon, A. M., & Schofield, G. (2019). Importance of machine learning for enhancing ecological studies using information-rich imagery. *Endangered Species Research, 39*, 91–104.

Gonzalez, Luis, Montes, Glen, Puig, Eduard, Johnson, Sandra, Mengersen, Kerrie, & Gaston, Kevin. (2016). Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors, 16*(1), 97. https://doi.org/10.3390/s16010097

Hasegawa, T., & Nakano, D. (2024). Robust fish recognition using foundation models toward automatic fish resource management. *Journal of Marine Science and Engineering, 12*(3), 488.

Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X., &, et al. (2024). Spectralgpt: Spectral remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 3*, 1.

John, B. A., Nelson, B., Sheikh, H. I., Cheung, S., Wardiatno, Y., Dash, B. P., Tsuchiya, K., Iwasaki, Y., & Pati, S. (2018). A review on fisheries and conservation status of asian horseshoe crabs. *Biodiversity and conservation, 27*, 3573–3598.

Johnston, D. W. (2019). Unoccupied aircraft systems in marine science and conservation. *Annual review of marine science, 11*(1), 439–463.

John, A., Shin, P. K., Botton, M. L., Gauvry, G., Cheung, S., & Laurie, K. (2021). Conservation of Asian horseshoe crabs on spotlight. *Biodiversity and Conservation, 30*, 253–256.

John, B. A., Sihin, S., Kamaruzzaman, B., Jalal, K., & Zaleha, K. (2012). Sediment profiling of the nesting grounds of horseshoe crabs at east peninsular malaysia. *International Journal of Biology, 4*(2), 159.

Kwan, B. K., Hsieh, H.-L., Cheung, S. G., & Shin, P. K. (2016). Present population and habitat status of potentially threatened Asian horseshoe crabs Tachypleus tridentatus and carcinoscorpius rotundicauda in hong kong: A proposal for marine protected areas. *Biodiversity and Conservation, 25*, 673–692.

Koyama, A., Hirata, T., Kawahara, Y., Iyooka, H., Kubozono, H., Onikura, N., Itaya, S., & Minagawa, T. (2020). Habitat suitability maps for juvenile tri-spine horseshoe crabs in Japanese intertidal zones: A model approach using unmanned aerial vehicles and the structure from motion technique. *PLoS ONE, 15*(12), 0244494.

Lee, C.N.-W., & Morton, B. (2016). Changes in the distributions of juvenile horseshoe crabs (Arthropoda: Chelicerata)(2002–2014) related to environmental perturbations at pak nai and ha pak nai, deep bay, hong kong sar, china. *Marine Pollution Bulletin, 108*(1–2), 134–146.

Lopez-Marcano, S., Brown, C. J., Sievers, M., & Connolly, R. M. (2021). The slow rise of technology: Computer vision techniques in fish population connectivity. *Aquatic Conservation: Marine and Freshwater Ecosystems, 31*(1), 210–217.

Li, J., Xu, M., Xiang, L., Chen, D., Zhuang, W., Yin, X., & Li, Z. (2024). Foundation models in smart agriculture: Basics, opportunities, and challenges. *Computers and Electronics in Agriculture, 222*, Article 109032.

Monteiro, J. G., Jiménez, J. L., Gizzi, F., Přikryl, P., Lefcheck, J. S., Santos, R. S., & Canning-Clode, J. (2021). Novel approach to enhance coastal habitat and biotope mapping with drone aerial imagery analysis. *Scientific Reports, 11*(1), 574.

Mou, C., Liang, A., Hu, C., Meng, F., Han, B., & Xu, F. (2023). Monitoring endangered and rare wildlife in the field: A foundation deep learning model integrating human knowledge for incremental recognition with few data and low cost. *Animals, 13*(20), 3168.

Morera, A. (2024). Foundation models in shaping the future of ecology. *Ecological Informatics, 130*, 102545.

Nagiewicz, S. D., Capone, V., Straub, P. F., Evert, S. P., & Robinson, N. (2022). Assessing the use of acoustic sampling for locating horseshoe crabs during migrations using side scan sonar. *International horseshoe crab conservation and research efforts: 2007–2020: Conservation of horseshoe crabs species globally* (pp. 453–468). Springer.

Nordlund, L. M., Torre-Castro, M., Erlandsson, J., Conand, C., Muthiga, N., Jiddawi, N., & Gullström, M. (2014). Intertidal zone management in the western indian ocean: Assessing current status and future possibilities using expert opinions. *Ambio, 43*, 1006–1019.

Pennings, S. C., Grant, M.-B., & Bertness, M. D. (2005). Plant zonation in low-latitude salt marshes: Disentangling the roles of flooding, salinity and competition. *Journal of Ecology, 10*, 159–167.

Rees, A. F., Avens, L., Ballorain, K., Bevan, E., Broderick, A. C., Carthy, R. R., Christianen, M. J., Duclos, G., Heithaus, M. R., Johnston, D. W., et al. (2018). The potential of unmanned aerial systems for sea turtle research and conservation: A review and future directions. *Endangered Species Research, 35*, 81–100.

Raffaelli, D. (1996). *Intertidal Ecology*. Chapman & Hall.

Ravoor, P. C., & Sudarshan, T. (2020). Deep learning methods for multi-species animal re-identification and tracking-a survey. *Computer Science Review, 38*, 100289.

Schneider, S., Taylor, G. W., Linquist, S., & Kremer, S. C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution, 10*(4), 461–470.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., Van Langevelde, F., Burghardt, T., et al. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications, 13*(1), 1–15.

Unger, S. D., Rollins, M., Dyer, N., & Payne, A. (2023). A walk on the beach: Exploring inaturalist observations of the Atlantic horseshoe crab, limulus polyphemus, as a tool for conservation. *Bulletin of Marine Science, 99*(4), 539–546.

Wang, S., Cheng, N., & Hu, Y. (2025). Comprehensive environmental monitoring system for industrial and mining enterprises using multimodal deep learning and clip model. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3533537

Wang, C.-C., Huang, S.-L., Wang, X., Xu, P., Huang, X., Liao, Y., Xie, X., & Kwan, K. Y. (2019). Conserving the understudied invertebrates: A call for a systematic monitoring protocol for Asian horseshoe crabs in nursery habitats. *Endangered Species Research, 40*, 369–373.

Wang, C.-C., Kwan, K. Y., Shin, P. K., Cheung, S. G., Itaya, S., Iwasaki, Y., Cai, L., Mohamad, F., Fozi, N. F., Zauki, N. A. M., et al. (2020). Future of Asian horseshoe crab conservation under explicit baseline gaps: A global perspective. *Global Ecology and Conservation, 24*, 01373.

Wisnewski, J., & Tanacredi, J. (2022). An assessment of horseshoe crab breeding habitat and trends on long island, New York, 2003–2017. *International horseshoe crab conservation and research efforts: 2007–2020: Conservation of horseshoe crabs species globally* (pp. 261–311). Springer.

Wang, X., Zhang, R., Kong, T., Li, L., & Shen, C. (2020). Solov2: Dynamic and fast instance segmentation. *Advances in Neural Information Processing Systems, 33*, 17721–17732.

Zhou, H., & Morton, B. (2004). The diets of juvenile horseshoe crabs, tachypleus tridentatus and carcinoscorpius rotundicauda (xiphosura), from nursery beaches proposed for conservation in hong kong. *Journal of Natural History, 38*(15), 1915–1925.

Zhang, M., Yang, B., Hu, X., Gong, J., & Zhang, Z. (2024). Foundation model for generalist remote sensing intelligence: Potentials and prospects. *Science Bulletin, 69*(23), 3652–3656.

Chen, T., Lu, A., Zhu, L., Ding, C., Yu, C., Ji, D., Li, Z., Sun, L., Mao, P., & Zang, Y. Sam2-adapter. (2024). Evaluating & adapting segment anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more. arXiv preprint arXiv:2408.04579

Doherty, K., Gurinas, M.A., Samsoe, E., Casper, C., Larkin, B.G., Ramsey, P.W., Trabucco, B., & Salakhutdinov, R. (2024). Leveraging foundation models for data-limited ecological applications. In: Neurips 2024 Workshop Foundation Models for Science: Progress, Opportunities, and Challenges. https://openreview.net/forum?id=s2OPqmsXMI

Fabian, Z., Miao, Z., Li, C., Zhang, Y., Liu, Z., Hernández, A., Montes-Rojas, A., Escucha, R., Siabatto, L., Link, A., & et al. Multimodal foundation models for zero-shot animal species recognition in camera trap images. arXiv preprint arXiv:2311.01064 (2023)

ICHO: Hong Kong Intangible Cultural Heritage Database (2024). https://www.hkichdb.gov.hk/en/index.html Accessed Dec 6, 2024

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., & *et al.* Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026 (2023)

Laurie, K., Chen, C., Cheung, S., Do, V., Hsieh, H., John, A., Mohamad, F., Seino, S., Nishida, S., Shin, P., & et al. Tachypleus tridentatus (errata version published in 2019). The IUCN Red List of Threatened Species 2019: e. T21309A149768986 (2019)

Lu, S., Guo, J., Zimmer-Dauphinee, J.R., Nieusma, J.M., Wang, X., VanValkenburgh, P., Wernke, S.A., & Huo, Y. (2024). Ai foundation models in remote sensing: A survey. arXiv preprint arXiv:2408.03464

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., & et al. Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023)

Praveena, K., Anandhi, R., Gupta, S., Jain, A., Kumar, A., & Saud, A.M. Application of zero-shot learning in computer vision for biodiversity conservation through species identification and tracking. In 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies, pp. 1–6 (2024). IEEE

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., & et al. Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458

Watanabe, S., Oyamada, S., Mizuta, K., Azumakawa, K., Morinobu, S., spsampsps Souji, N. Activity rhythm of the tri-spine horseshoe crab tachypleus tridentatus in the seto inland sea, western japan, monitored with acceleration data-loggers. In: International Horseshoe Crab Conservation and Research Efforts: 2007-2020: Conservation of Horseshoe Crabs Species Globally, pp. 107–121. Springer, Cham (2022)

Xiao, A., Xuan, W., Wang, J., Huang, J., Tao, D., Lu, S., & Yokoya, N. (2024). Foundation models for remote sensing and earth observation: A survey. arXiv preprint arXiv:2410.16602

Authors and Affiliations

Jihan Zhang[1] · Mingqiao Han[1] · K. H. Laurie[2] · Benyun Zhao[1] · Lei Lei[1] · Xi Chen[1] · Hon Chi Judy Wan[3] · Siu Gin Cheung[4] · Wenxing Hong[5] · Ben M. Chen[1]

✉ Mingqiao Han
1155230078@link.cuhk.edu.hk

✉ Xi Chen
xichen002@cuhk.edu.hk

Jihan Zhang
1155139089@link.cuhk.edu.hk

K. H. Laurie
horseshoecrab@ymail.com

Benyun Zhao
1155145791@link.cuhk.edu.hk

Lei Lei
leilei001@cuhk.edu.hk

Hon Chi Judy Wan
judy.hcj.wan@oceanpark.com.hk

Siu Gin Cheung
bhsgche@cityu.edu.hk

Wenxing Hong
hwx@xmu.edu.cn

Ben M. Chen
bmchen@cuhk.edu.hk

1 Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, N.T, Hong Kong, China

2 IUCN SSC Horseshoe Crab Specialist Group, Hong Kong, China

3 Ocean Park Conservation Foundation, Ocean Park, Aberdeen, Hong Kong, China

4 Department of Chemistry, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China

5 School of Aerospace Engineering, Xiamen University, Xiamen, China