Supplementary material 1. Details of the working steps followed in the process of fauna identification from images.

1. Non-expert annotation processing

First, we delimited each faunal observation as an individual to be identified within images (segmentation). In the context of biodiversity metrics comparison between physical sampling and imagery, we conserved all images annotated for analyses, even some poor-quality images (i.e., too dark or light, too low or high relative to the sea bottom) to compare all the taxonomic groups and identification ranks accessible from images compare to samples. In PNG area, the pool of poor quality images represents 1.6% of the dataset while in Mayotte, it represents 7%.

Second, each individual was assigned a label corresponding to a high taxonomic classification rank from phylum to class. These classifications and identifications relied on objective observations of morphological features that are diagnostic of each high-ranking taxonomic group and do not require advanced taxonomic skills. Non-recognizable individuals, even at a phylum rank were dismissed and attributed an "Animalia cetera" label.

To facilitate more precise identifications in collaboration with taxonomists, we delimited preliminary morphotypes for some taxonomic groups based on observations of similar morphological features between individuals to be identified from images and photos of collected specimens.

2. Objective identification processing

In step three, in collaboration with taxonomists, we evaluated the morphotypes/high-rank taxonomic groups on the basis of objective morphological characters that can be observed in images. These characters correspond to diagnostic characters (i.e., characters differentiating taxa), allowing the identification of individuals at a taxonomic rank N with certainty. Therefore, each set of individuals sharing the same diagnostic characters was defined as a photo-taxon (i.e., a set of individuals classified into the same taxonomic entity) and then assigned to a taxonomic rank N (e.g., genus rank).

The photo-taxon concept was originally proposed to divide organisms in images into operational taxonomic units (OTU) (used by Williams et al., 2015 and defined in Althaus et al., 2009). Like morphospecies, these OTUs may be difficult to associate with a valid name, particularly at the species level. The photo-taxa are thus associated with different taxonomic ranks.

Individual organisms whose images showed blurred characters that did not allow an assessment of their morphological characters remained labelled at a high taxonomic rank only (e.g., phylum, class).

In order to specify the identification of a photo-taxon at a lower rank, some poor-quality images (blurriness, view angle, missing characters) had to been set aside. In this case, we added the qualifier "indeterminabilis" to the rank assigned to the photo-taxon, which is consistent with open nomenclature proposed by Sigovini et al. (2016) and more recently adapted by Horton et al. (2021) for image-based identification. The remaining good quality images of the photo-taxon were used to precise its identification rank.

3. Contextual identification processing

In step four, we proceeded with a contextual identification, meaning that photo-taxa assigned to a reliable taxonomic rank N in images were specified at a lower taxonomic rank (N+1) from contextual elements. These elements are based on the knowledge of taxa collected in the study area and potentially similar to the specimens to be identified in images. They reflect the experts' knowledge and data from the literature, and consider: (1) the morphological characters of collected species/taxa that we can compare with individuals to be identified in images; (2) the ecology of the species/taxa in the study area (depth range, suitable substrate, preferred prey, species association, position relative to substrate, etc.); (3) character variability (intraspecific polymorphism, interspecific convergence of characters) and (4) species distribution and diversity level (e.g., number of known species within a genus, number of known genus within a family, etc.).

For each photo-taxon identified at a given taxonomic rank N, we computed the list of the known taxa attributed to this name (e.g., if the photo-taxon is attributed to a genus name, then we listed all the species belonging to this genus, in the physical samples analyzed in this study and from physical sampling undertaken during previous cruises in the region). For each photo-taxon, the set of identified physical specimens allowed us to evaluate the characters that might be used on the images to refine the identification of the photo-taxon. This step resulted in defining a sequence of suitable diagnostic characters for image-based identification. Potential species whose diagnostic characters did not match those in the images of the photo-taxon to be identified were eliminated from the list.

By elimination, we were able to refine the identification of photo-taxa to lower taxonomic levels. Either photo-taxa were assigned to a valid name of rank N+1 (corresponding to the name of the most likely taxon) or it was not possible to assign a valid species name or morphospecies. We delimited morphospecies when, for a photo-taxon identified at a reliable rank N, the observation among individuals belonging to this photo-taxon allowed the differentiation of two or more subsets, based on the observation of characters that are known to be not shared within a single species. However, most often, observable characters were not sufficient to attribute a valid name to these species.

4. Identification key construction

We selected good-quality images that were illustrative of the identified photo-taxon at rank N or N+1 (when finer identification was possible). This set of identified images was considered as a reference in the identification process and is referred to as photo-types. The catalog of photo-types could be easily built with the *Annotation catalog* option proposed by BIIGLE 2.0 allowing a mosaic view of the different photo-taxa for each taxonomic entity (label).

Lastly, we compiled the diagnostic characters used to identify photo-taxa, and classified them into distinct modalities to construct identification keys. Based on photo-types and the knowledge of diagnostic characters, it was possible to identify the remaining photo-taxa.