Supplementary material

S1. High-resolution version of Figures 2 and 4

High-resolution versions of Figures 2 and 4 are provided below, allowing clearer visualisation of fine-scale details. For better visualisation, these figures are presented in a larger format than the main paper. Figs. 1 to 4 correspond to Figure 2 in the main paper, while Figs. 5 to 8 correspond to Figure 4.

Figs. 7 and 8 show a zoomed-in view of the UAV orthophoto and the corresponding predicted segmentation mask for the *Trou d'eau* lagoon, respectively. In Fig. 8, black circles highlight spurious patches of different corals predicted by the model, which are not present in the orthophoto. Despite these errors, the overall quality of the segmentation mask is encouraging, demonstrating the model's ability to capture the spatial distribution of coral morphotypes and habitats across the reef area.

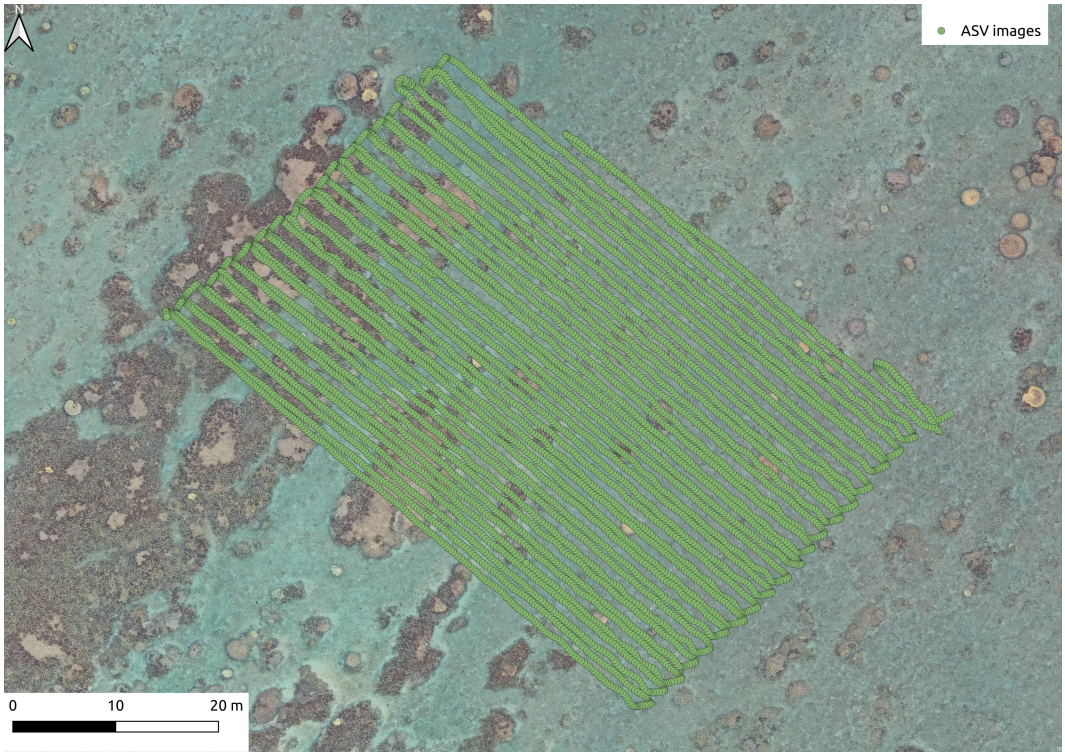


Figure 1. Example of the rasterization process on an ASV data collection event in the *Trou d'eau* lagoon in Reunion Island: GPS coordinates of ASV images along the session. The high density of images allows for fine-scale spatial interpolation.

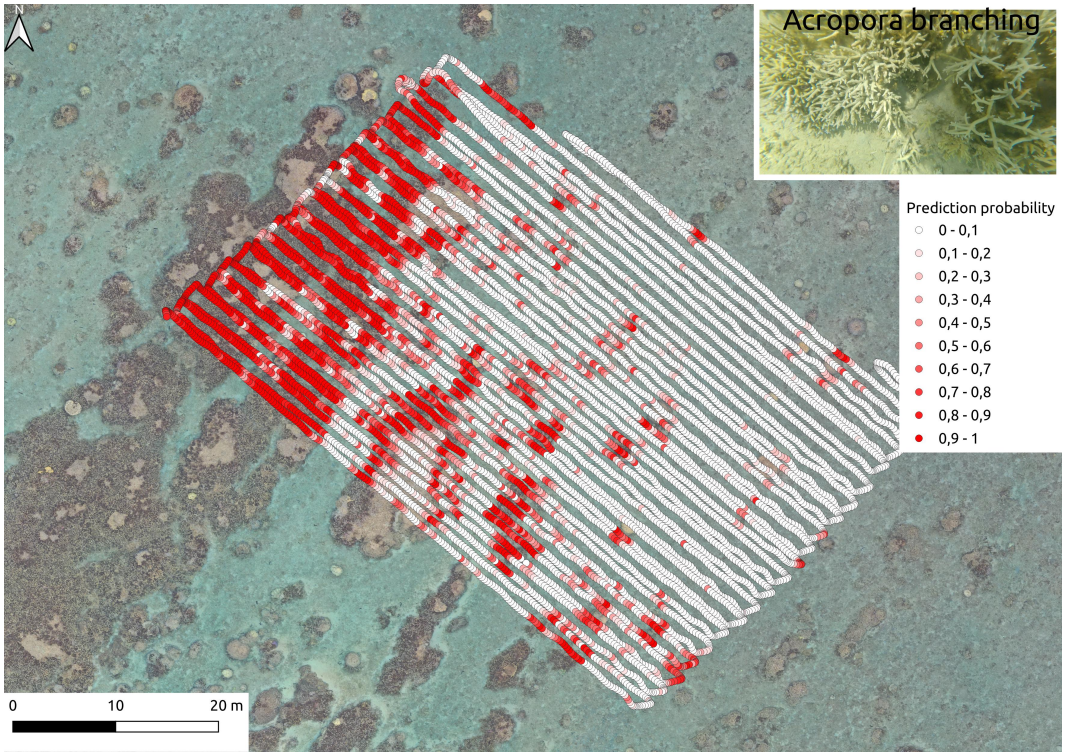


Figure 2. Example of the rasterization process on an ASV data collection event in the *Trou d'eau* lagoon in Reunion Island: pointwise predictions of *Acropora Branching* presence in each ASV image.

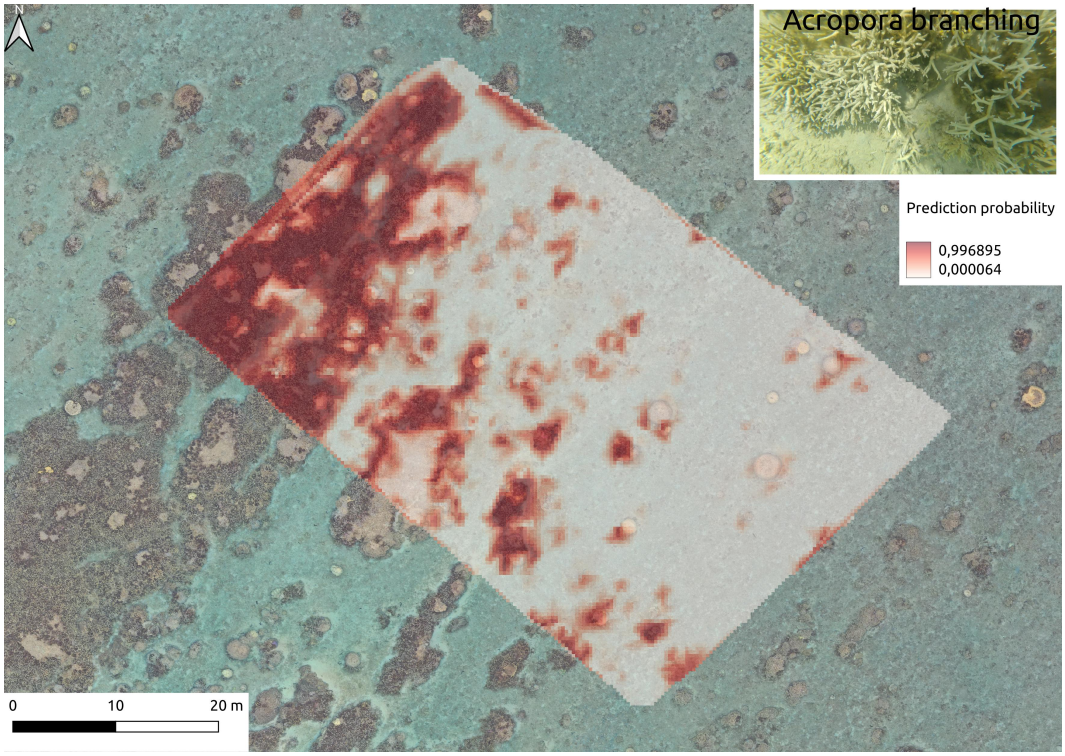


Figure 3. Example of the rasterization process on an ASV data collection event in the *Trou d'eau* lagoon in Reunion Island: interpolated raster representing the probability of *Acropora Branching* presence at each grid point.

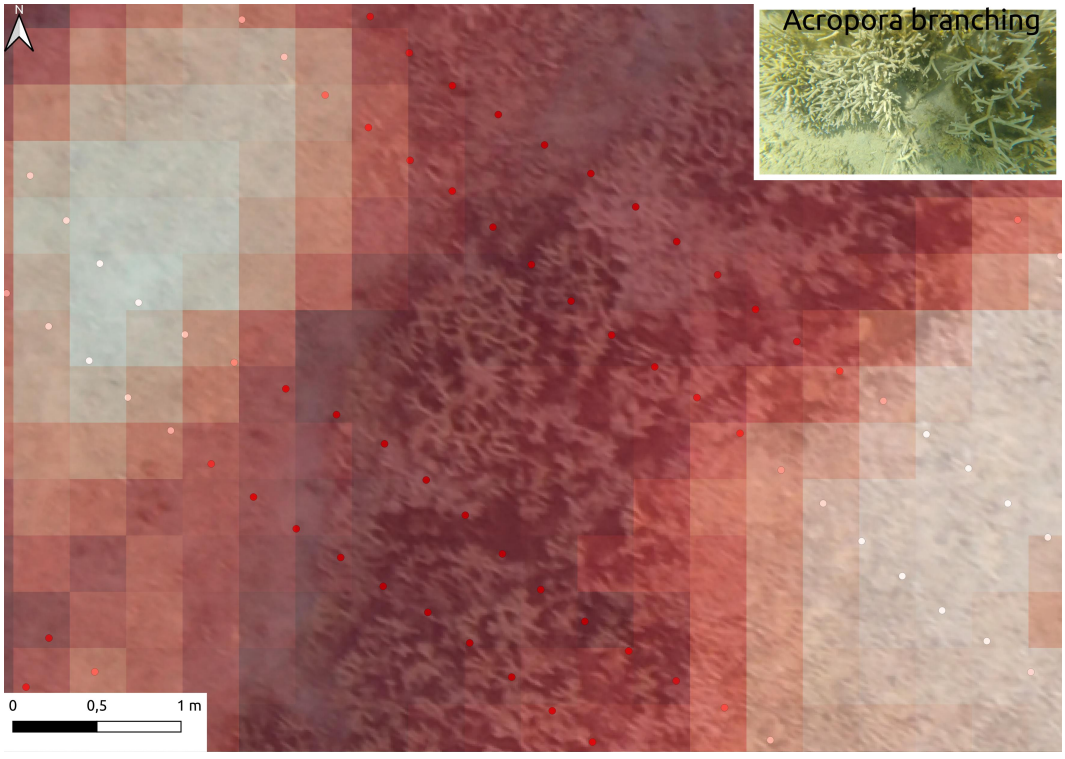


Figure 4. Example of the rasterization process on an ASV data collection event in the *Trou d'eau* lagoon in Reunion Island: zoomed-in view of the raster and corresponding underwater prediction points. The raster provides a continuous spatial representation of coral morphotype distributions, enabling the generation of rough segmentation masks that can be used for UAV model training.

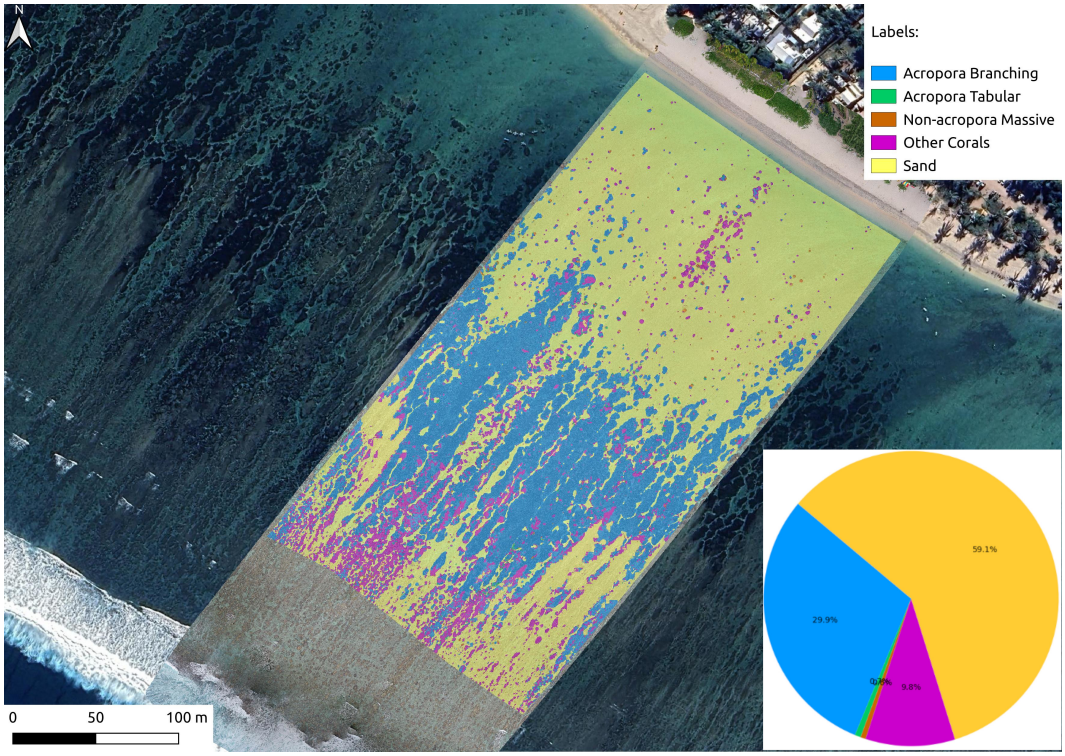


Figure 5. Results of the UAV model applied to the entire orthophoto of the *Trou d'eau* lagoon. The model predicts the presence of coral morphotypes and habitats across the entire reef area. A pie chart is used to quantify the predicted habitat composition.

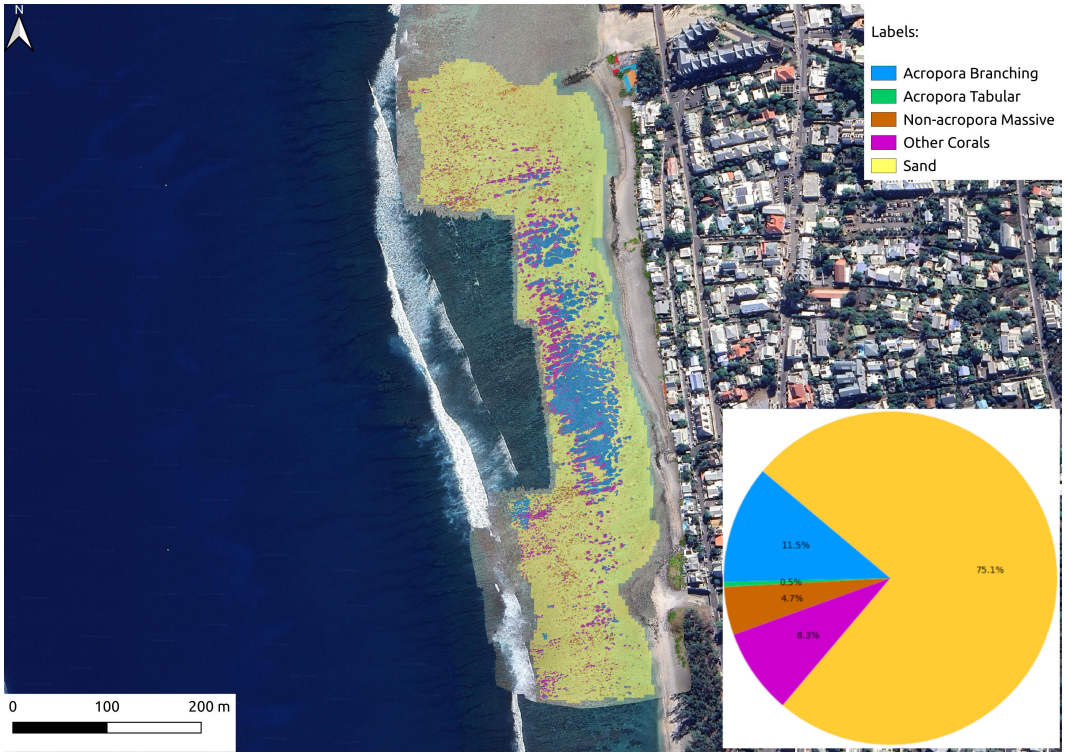


Figure 6. Results of the UAV model applied to the entire orthophoto of the *Saint-Leu* lagoon. The model predicts the presence of coral morphotypes and habitats across the entire reef area. A pie chart is used to quantify the predicted habitat composition.

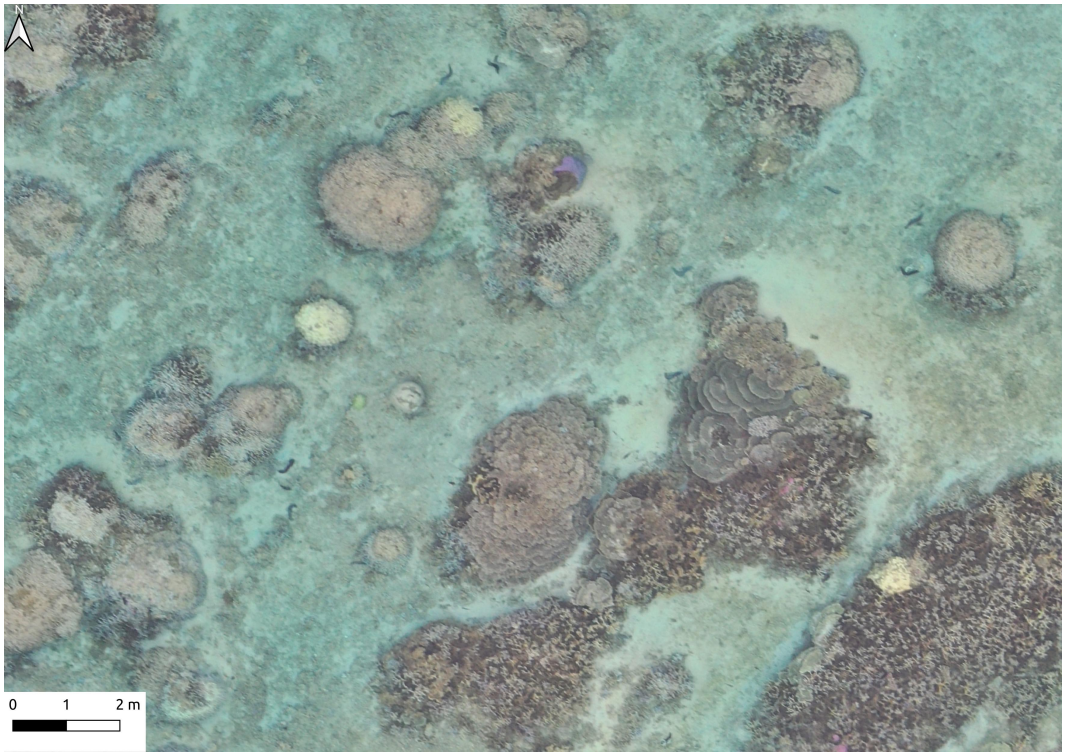


Figure 7. Zoomed-in view of the UAV orthophoto of the *Trou d'eau* lagoon.

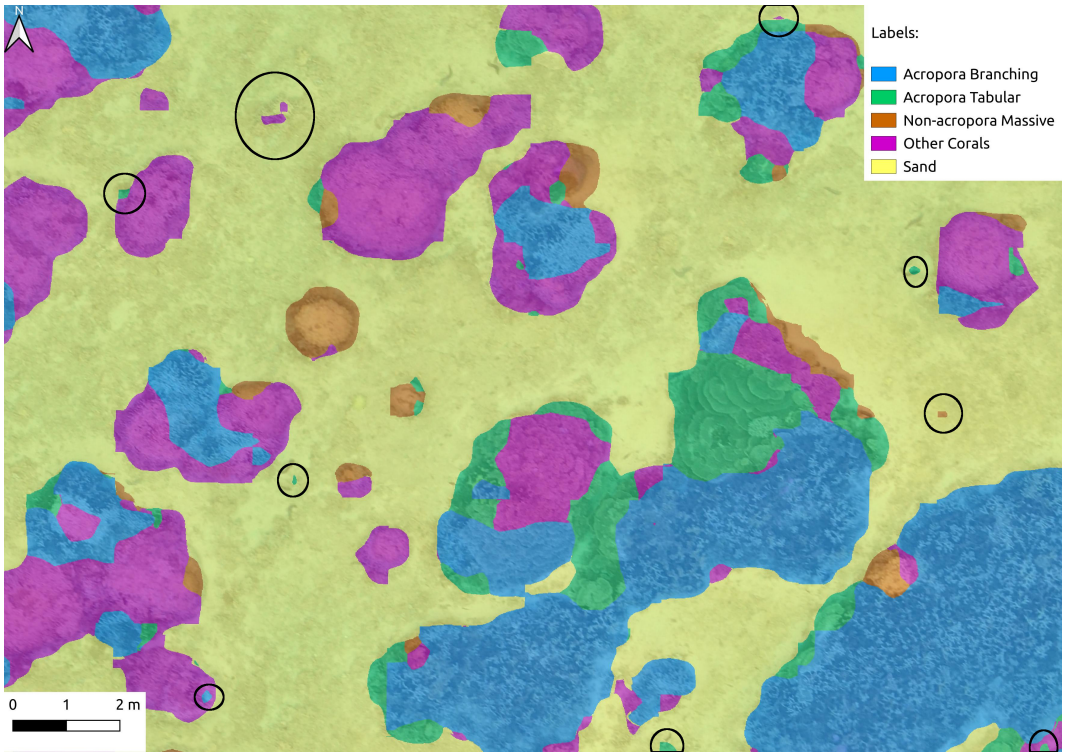
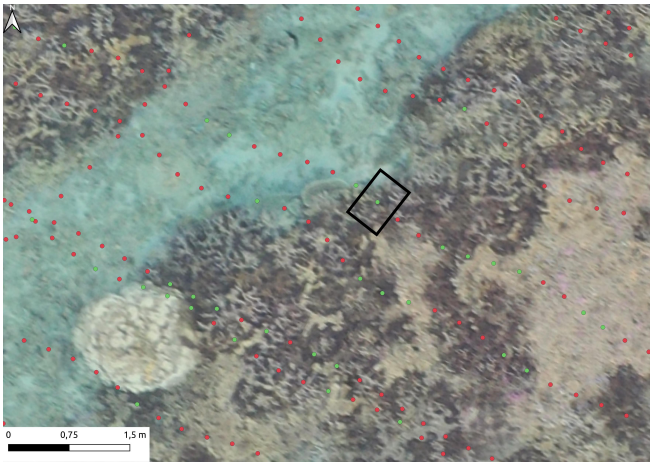


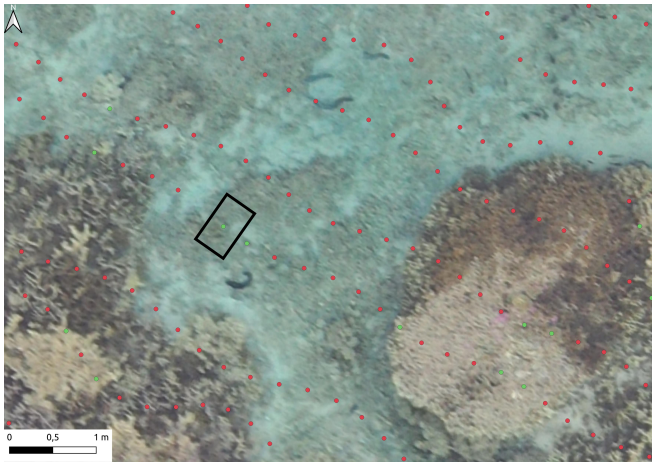
Figure 8. Corresponding predicted segmentation mask for the zoomed-in area of the *Trou d'eau* lagoon orthophoto, showing the spatial distribution of coral morphotypes and habitats. Black circles highlight spurious patches of different corals predicted by the model, which are not present in the orthophoto.

S2. Examples of classes not reliably distinguishable from aerial imagery

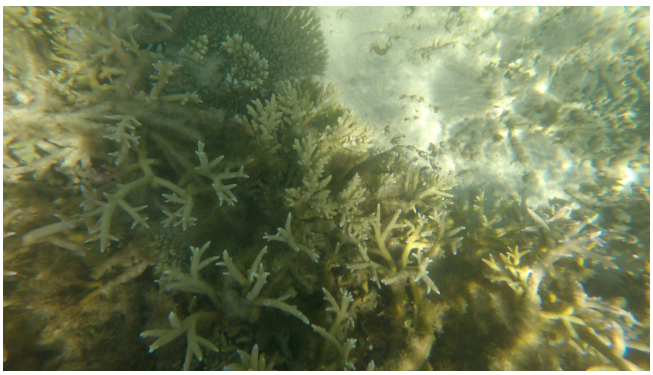
Some classes used in the underwater teacher model cannot be confidently identified in UAV orthophotos due to spatial resolution constraints or spectral ambiguity. Fig. 9 shows two illustrative examples comparing aerial and underwater views of the same locations, highlighting this limitation.



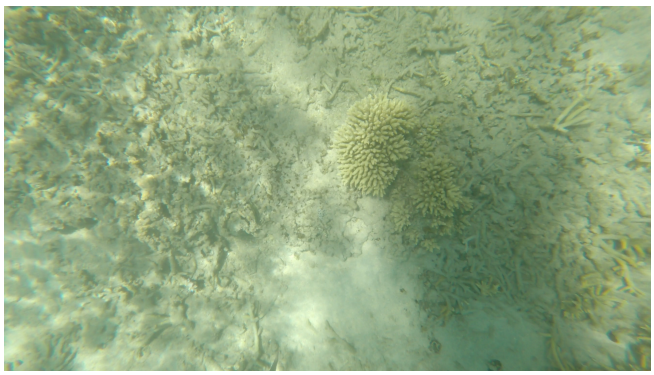
(a) Aerial view of a reef zone with predicted *Algal Assemblage* presence. Green dots indicate ASV image locations where the underwater *DinoVdeau* model predicted the presence of *Algal Assemblage*; red dots indicate absence. The black rectangle highlights the location of the underwater image shown in Fig. 9c, where a dense algal cover is confirmed.



(b) Aerial view of a reef zone with predicted *Acropora Digitate* presence. Green dots indicate ASV image locations where the underwater *DinoVdeau* model predicted the presence of *Acropora Digitate*; red dots indicate absence. The black rectangle highlights the location of the underwater image shown in Fig. 9d, where a small colony of *Acropora Digitate* is confirmed.



(c) Underwater image taken by an ASV camera showing a dense algal assemblage, confirming *DinoVdeau* prediction at this location.



(d) Underwater image showing a small colony of *Acropora Digitate*, which is ecologically relevant but not distinguishable in the corresponding aerial orthophoto due to its small size and low contrast.

Figure 9. Comparison of aerial orthophotos (top) and corresponding underwater images (bottom) for two example locations. These examples illustrate how some benthic classes, such as *Algal Assemblage* (left) and small coral morphotypes like *Acropora Digitate* (right), can be clearly identified in underwater imagery but remain indistinct in aerial views due to resolution and colour similarity limitations.

Fig. 9a and Fig. 9b show two examples of aerial orthophotos with green and red dots indicating the presence or absence of *Algal Assemblage* and *Acropora Digitate* predicted by the underwater *DinoVdeau* model on the corresponding ASV images. The black rectangles highlight the locations of the underwater images shown in Fig. 9c and Fig. 9d, where the presence of *Algal Assemblage* and *Acropora Digitate* is confirmed. These examples illustrate how some benthic classes, such as *Algal Assemblage* (left) and small coral morphotypes like *Acropora Digitate* (right), can be clearly identified in underwater imagery but remain indistinct in aerial views due to resolution and colour similarity limitations.

Other classes, although ecologically important and distinguishable on UAV orthophotos like seagrass species *Syringodium isoetifolium* and *Thalassodendron ciliatum* were excluded. This decision was based on the limited number of presence predictions available from the underwater model, which would have increased label noise during training. Fig. 10 presents the relative abundance of each benthic class predicted in the underwater imagery.

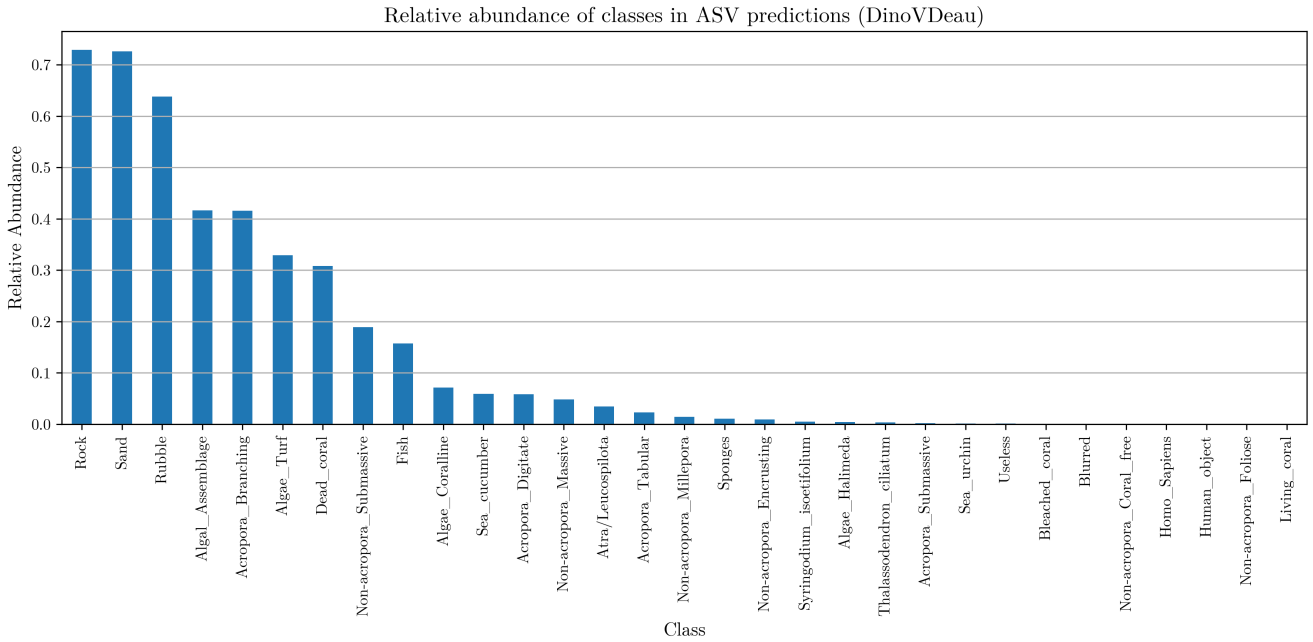


Figure 10. Relative abundance of benthic classes predicted in the underwater imagery by *DinoVdeau* model across the six training ASV sessions.

Hard-substrate classes like *Rock*, *Sand* and *Rubble* dominate the distribution but, as mentioned earlier, only *Sand* was retained in the UAV model. *Rock* and *Rubble* were excluded because they appear too similar to *Sand* in UAV imagery at the given resolution. Additionally, their frequent co-occurrence in underwater predictions would have introduced significant label noise during the generation of coarse annotations. The coral morphotypes *Acropora Branching*, *Acropora Tabular* and *Non-acropora Massive* were included in the UAV model based on their relatively high abundance and distinctive appearance in UAV orthophotos. Lastly, *Other Corals* represents a heterogeneous group of morphotypes that lack consistent visual cues in aerial imagery and are therefore grouped together to reduce annotation ambiguity.

It is also important to note that the relative abundance values shown in Fig. 10 do not sum to 1. This is because the underwater model *DinoVdeau* is a multi-label image classification model, allowing multiple classes to be predicted for a single image, reflecting the natural co-occurrence of benthic features in reef environments.

S3. WSSS loss function

When dealing with WSSS the choice of the loss function can have a significant impact on the model performance. Indeed during training, the loss function is used at each iteration to compute the error between the predicted segmentation mask and the ground truth mask. Since the ground-truth masks are generated from weak annotations, they contain noise and inaccuracies, which must be taken into account when choosing the loss function. For the first training of Segformer model, we tried four different loss functions: boundary loss, dice loss, cross entropy (CE) loss and focal loss.

The comparison between the coarse annotation mask and the prediction masks obtained by training the model shows substantial differences between the four different loss functions (Figure 11). The CE loss and focal loss are pixel-level losses that focus on individual pixels [1], making them sensitive to label noise and class imbalance. This is reflected in the predictions, that often do not recognise under-represented classes such as *Acropora Tabular* and *Non-acropora Massive*. Boundary level losses, like the boundary loss, are designed to focus on the edges of objects, which can be useful for fine-grained segmentation tasks. For WSSS however, boundary losses are too sensitive to noise in the annotations and this is clearly reflected in the predictions in Figure 11, which exhibits high pixelation effects.

Finally dice loss [2], although sometimes produces noisier masks than the ones obtained with other losses, is the only one that correctly handled the class imbalance. In fact, the dice loss is a region-based loss that evaluates the overlap between predicted and annotated regions as a whole, rather than focusing on individual pixels or precise edges. This makes the loss robust to class imbalance even if less sensitive to fine edge details or precise boundary alignment.

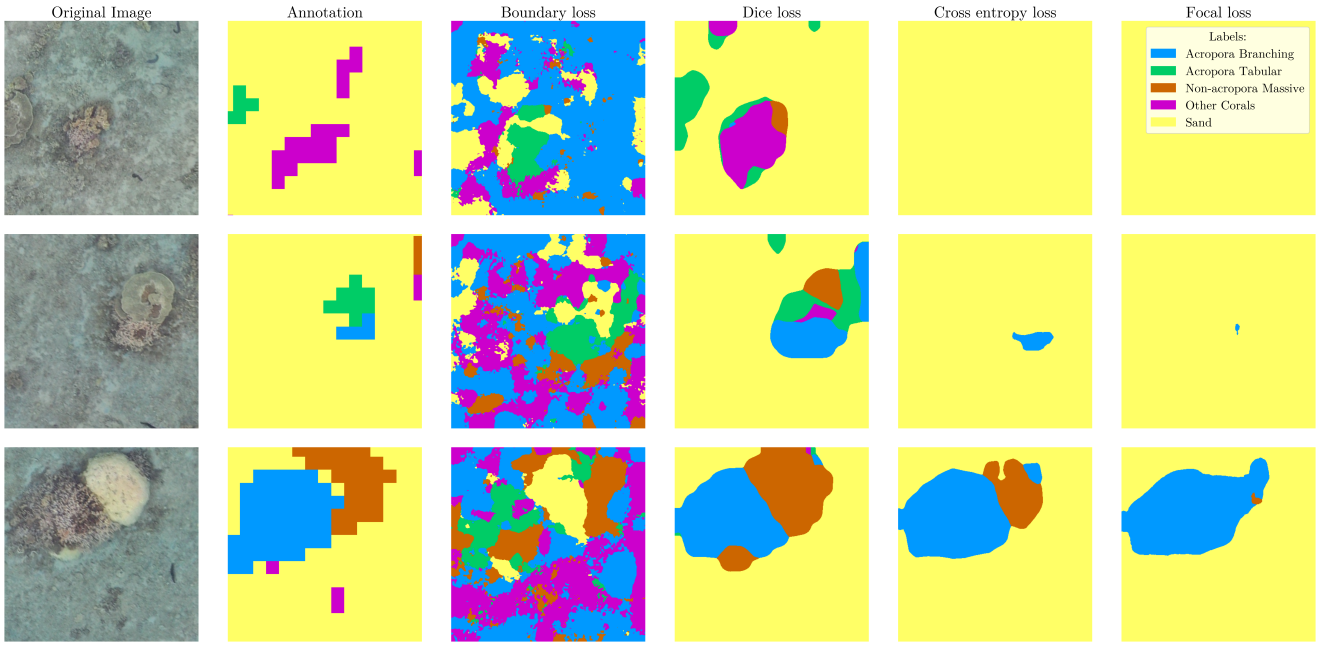


Figure 11. Comparison of four different loss functions for WSSS training with coarse annotations. The first column shows the original UAV tile, followed by the coarse annotation used for training. The remaining columns display the predicted segmentation masks obtained using boundary loss, dice loss, cross-entropy loss and focal loss, respectively.

S4. Performance and integration of the *Sea Cucumber* class

To evaluate the framework's capacity to incorporate new classes such as mobile species, we extended the segmentation model to include a new *Sea cucumber* class, using a small number of manual annotations generated via the Geo-SAM plugin. The model was retrained from scratch, using the same configuration and ASV-derived annotations as before, but with the manually added *Sea cucumber* masks overlaid in the training data.

To mitigate class imbalance, a weighted Dice loss was applied, assigning a higher penalty (seven times the weight of other classes) to errors involving *Sea cucumber*. The training process did not involve fine-tuning from previous weights to avoid bias toward earlier class distributions and to assess whether the inclusion of new categories interferes with performance on existing ones.

The model was tested on the same test zones as before, specifically the *Trou d'eau* and *Saint-Leu* lagoons, by adding the *Sea cucumber* class to the ground truth annotation masks. Model evaluation showed that the new class was detected with high accuracy:

- Pixel Accuracy: 0.7545
- IoU: 0.4817

demonstrating the model's ability to learn from a small number of manual annotations.

To assess whether introducing *Sea cucumber* impacted the model's performance on other classes, we compared standard metrics with those from the original 5-class version. As shown in Tab. 1, a decrease in overall accuracy (from 85.72% to 83.35%) and mean IoU (from 50.10% to 44.48%) was observed, indicating a modest reduction in segmentation performance after the addition of the new class.

Table 1. IoU per class, total accuracy and mean IoU with the extended 6-class model including *Sea cucumber*.

Zone	Accuracy	Mean IoU	Acropora B.	Acropora T.	Non-acro M.	Other Corals	Sand	Sea Cucumber
Total	**0.8335**	**0.4448**	**0.4638**	**0.1560**	**0.3905**	**0.2511**	**0.9254**	**0.4817**

The IoU scores for all coral-related classes declined: for *Other Corals* (from 0.4111 to 0.2511) and *Acropora Tabular* (from 0.2313 to 0.1560). Similarly, *Acropora Branching* and *Non-acropora Massive* saw moderate reductions (from 0.5282 to 0.4638 and 0.4095 to 0.3905, respectively). These drops suggest that the integration of a new, visually distinct class, such as *Sea cucumber*, may have introduced some degree of class interference, likely due to the high penalty weighting assigned to it in the loss function. Importantly, the performance on the *Sand* class remained stable (0.9249 vs 0.9254), suggesting that well-represented, visually dominant classes are more robust to such extensions.

Overall, while the model preserved general robustness and successfully learned to identify the new *Sea cucumber* class with reasonable accuracy (IoU = 0.4817), this resulted in a measurable decline in the precision of some coral classifications. These findings highlight the importance of considering potential trade-offs when expanding class taxonomies, particularly in class-imbalanced and visually complex environments.

References

[1] Reza Azad, Moein Heidary, Kadir Yilmaz, Michael Hüttemann, Sanaz Karimijafarbigloo, Yuli Wu, Anke Schmeink, and Dorit Merhof. Loss functions in the era of semantic segmentation: A survey and outlook. *arXiv preprint arXiv:2312.05391*, 2023. 8

[2] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017. 8