

A Validation Tool For Improving Semantic Segmentation of Complex Natural Structures

G. Pavoni^{1,2}, M. Corsini^{2,3} , M. Palma⁴ and R. Scopigno²

¹Dipartimento di Ingegneria dell'Informazione, Pisa, Italy

²ISTI-CNR, Pisa, Italy

³Università di Modena e Reggio Emilia, Modena, Italy

⁴DISVA, Università Politecnica delle Marche, Ancona, Italy

Abstract

The automatic recognition of natural structures is a challenging task in the supervised learning field. Complex morphologies are difficult to detect both from the networks, that may suffer from generalization issues, and from human operators, affecting the consistency of training datasets. The task of manual annotating biological structures is not comparable to a generic task of detecting an object (a car, a cat, or a flower) within an image. Biological structures are more similar to textures, and specimen borders exhibit intricate shapes. In this specific context, manual labelling is very sensitive to human error.

The interactive validation of the predictions is a valuable resource to improve the network performance and address the inaccuracy caused by the lack of annotation consistency of human operators reported in literature.

The proposed tool, inspired by the Yes/No Answer paradigm, integrates the semantic segmentation results coming from a CNN with the previous human labeling, allowing a more accurate annotation of thousands of instances in a short time. At the end of the validation, it is possible to obtain corrected statistics or export the integrated dataset and re-train the network.

CCS Concepts

• **Human-centered computing** → *Graphical user interfaces; Visual analytics*; • **Computing methodologies** → *Image segmentation*;

1. Introduction and Motivation

The increasing popularity of low-cost underwater video cameras has facilitated the collection of large visual datasets for underwater monitoring purposes. The automatic analysis of this massive amount of data remains an open issue, exemplified by the estimation that only the 1-2% of the acquired images is subsequently analyzed by an expert human operator [BEK*12].

Neural networks have been successfully used for the automatic classification of marine organisms. In particular, CNN have demonstrated to obtain good performance in the semantic segmentation of benthic communities [KBH18]. While speeding up the recognition step, the use of neural networks require the preparation of a large training dataset, done by experienced personnel, resulting in a very time-consuming process. Moreover, different studies underline the weak consistency of human annotations on coral reefs survey images. [NDOS03] compared *in situ* observations with image annotations reporting different accuracies in detecting different marine species (92.5% of accuracy in distinguishing *soft corals*). In [BER*15], authors measured the accuracy in distinguishing coral taxa between two groups of biologists; the “Hosts”,

familiar with coral classification on a specific area, and the “Visitors” experts in classifying corals from other geographical regions. Measured with Cohen’s kappa statistic, the “Hosts” reach an accuracy in distinguishing between local coral genera of about the 79.4% while the “Visitors” settles at 58.6%. It is very easy, even for experienced operators, to confuse a coral genera with other coral genera. However, distinguishing between corals remains a simpler task than distinguishing other marine organisms. The accuracy of a local operator in the distinction between Crustose Coralline Algae from images is just the 51.0%. Bejibom et al. [BER*15] estimate a very interesting fact, the *intra-annotator variability*, the agreement of a “Host” operator in classifying two times the same coral genera, is of about the 89.7%.

Recently, interactive tools to speed up the annotation process and/or improve the training of the networks have been extensively developed. Papadopoulos et al. [PUKF16] suggested to exploit the Yes/No paradigm for fast semi-automatic annotation of object using bounding boxes. In another work, Papadopoulos et al. [PUKF17] demonstrated that click on the extreme part of an object it is more efficient than indicating it with a bounding box. Cui et al. [CZLB16] proposed a tool to improve fine-grained categoriza-

tion. This tool, after an automatic pre-classification, ask the user to validate the results giving a Yes/No answer. These answers are then used to train a deep learning metric specifically designed to exploit the human indications. RNN++ [ALKF18], that is an evolution of the RNN tool [CKUF17] for object segmentation, uses a recursive neural network to learn the border of the image to segment during the manual polygonal annotation.

In this paper we do not propose an annotation tool but an interactive *validation* tool aimed at improving the performances of a semantic segmentation network together with the human annotation accuracy. Validate the results of the segmentation of natural structures is important, because, in general, are used to calculate accurate metrics. For example, in the case here presented metrics related to the spatial analysis of populations (such as abundance or coverage) are of particular importance. The validation tool takes inspiration from the Yes-No Answer paradigm to speed up the validation process. The main idea is that the user has to choose the best segmentation between the initial human annotation and the network predictions. At the end of the validation it is possible to export a new dataset, corrected of some of the initial errors.

2. Case Study

We tested our validation tool on a 150×50 meters wide ortho-mosaic depicting a portion of the Mozambique coral reef already studied in Palma et al. [PCPC17]. This area contains various coral species (see Fig. 1), among them the *Soft Coral Digitate* (*SCD*) class was manually labeled drawing polygons around corals inside a GIS software (QGIS). The polygonal annotation of ortho-mosaics is a raising trend among biologists, well suited for semantic segmentation tasks.

The *SCD* class, which shows a large intraspecific morphological variance, covers approximately just the 6.4% of the seabed. From the perspective of an intelligent monitoring system distinguish between a *coral taxa / other* is a tricky problem. The 'other' class contains elements that are poor in features, such as sand, but also other corals classes morphologically similar to the monitored one. This can lead easily to false positive predictions. Note that this problem shares some similarities with fine-grained recognition, presenting a large intra-variance class and a low inter-variance class.

In our study, the lower portion of the map (dashed in Fig. 1) was selected to be the test area; all the rest was used to feed a Bayesian Segnet [KBC15], a standard CNN architecture for pixel-wise classification problems. The network achieved the 95.1% of accuracy in classifying pixels of the *SCD* class, with the following confusion matrix: $TP = 70.3\%$, $FP = 3.5\%$, $FN = 2.97\%$, $TN = 96.5\%$. Positive predictions of the *SCD* class were later re-mapped on the test area and validated by the same biologist into QGIS. Results are visible in Fig. 2. Purple corals indicate those ones classified coherently by the biologist and the network. Red polygons were initially labelled as *SCD*, correctly predicted by the CNN, but later evaluated as belonging to the 'other' class by the biologist (*false positive in original labelling*). On the contrary, yellow polygons were initially annotated as belonging to the 'other' class, incorrectly segmented by the CNN, but later recognized by the human operator as *SDC* (*false negative in original labelling*). The classification errors

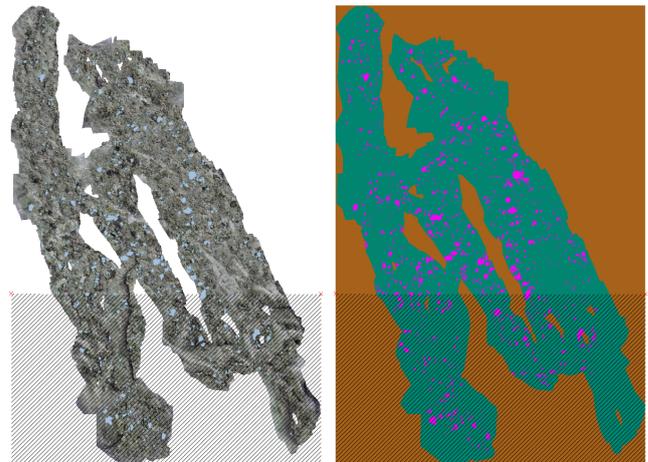


Figure 1: The Mozambique dataset. Ortho-mosaic maps depicting the part of the Mozambique Coral Reef (on the left), Soft Coral Digitate Class is highlighted in light blue. Corresponding labelled map (on the left). The dashed part correspond to the network's test area.

introduced by the biologist agree with the literature's statistics for this specific task [BER*15]. We estimated that:

- Red pixels are ~ 423.0000 , about the 16% of the total amount of the true *SCD* class pixels in the test area. There are few false positive instance in original labelling, but, given the morphological similarity between different classes, this can lead to significant errors.
- Yellow pixels are ~ 730.000 , about the 37% of the total amount of the true *SCD* class pixels in the test area. There are a considerable quantity of corals initially unnoticed, but recognized when proposed as positive prediction by the network. Again, from a learning perspective, all the forgotten data wrongly classified in the training could be serious issue.
- In some cases (see Fig. 4, top-right image), the annotation is consistent at the class level but not at the coral contour level. This is a minor threat which we must take into account since we are considering per-pixel classification errors.

The 16% of the total amount of pixels were wrongly classified by the human operator in the test area during the original labelling, i.e. the 43% of the *SCD* class. Ideally, assuming the total absence of errors in the training area, this would change the positive prediction of the confusion matrix to $TP = 73.0\%$, $FP = 3.4\%$.

3. Tool description

The Mozambique's coral reef case study allows us to better understand, quantitatively and qualitatively, the classification errors done by human annotators. These errors represent an additional source of uncertainty in the challenging task of distinguishing between similar coral taxa (as demonstrated in [KBH18]).

The biologist needed approximately 25 working hours to re-assign every predicted coral to its correct class. The validation

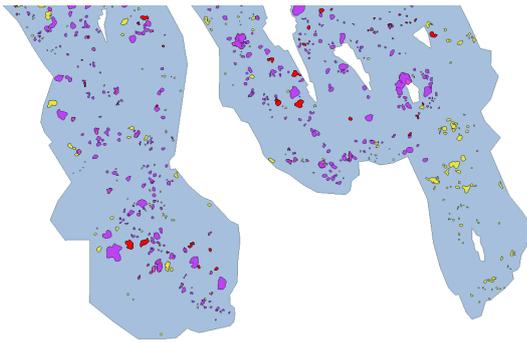


Figure 2: Comparison between the original polygonal annotation of the Soft Coral Digitate class and the validation.

step in a GIS software, which is a typical choice, is very time-consuming, does not offer immediate feedback on the predictions statistics and, finally, prepare a new training dataset is a labor-intensive process.

The proposed tool provides a user interface and a series of simple click-based operations specifically designed for the validation task and for this type of dataset. The user validates the existing annotations by performing three operations: *accepting* a positive prediction (new instances or instances segmented more accurately than labels), *discarding* an unwanted positive prediction (false positives or instance less accurately segmented than labels), and finally *copying* a label not detected by the network (false negative predictions). This validation method differentiates between the incorrectly-predicted corals and the missed ones: storing these information allows to perform measurement task with higher precision (e.g. area covered by the SCD), and the false positive and false negative predictions are additional information that can be used to re-train the network.

The interface is subdivided into three parts (see Fig. 3). The Comparison Panel offers a simultaneous view of the original labeling (on the left) and of the segmentation results (on the right), superimposed to the ortho-mosaic map. The two views are synchronized, and the transparency of both labelling can be controlled with a slider. This side-by-side visualization makes the comparison between the previous labeling and the output of the network easier, with respect to a layer-based visualization.

Since the ortho-photo map is about 600 Mpixels, we need a help in navigating it. The Navigation Panel offers a summary view of the whole map, showing subdivision in tiles: each tile can be accessed by simply clicking on it. The amount of validated tiles, as well as the remaining ones are visualized in the Control Panel. The Control Panel contains also all the commands to load/save and export the integrated annotations (for re-training) and the buttons related to the validation operations performed by the tool *Accept/Reject/Copy*.

The workflow is the following: the tool pre-process the human labels together with the ones given by the CNN and creates two sets of blobs (label blobs and segmentation blobs). Then, the expert choose a tile to work on, select a segmented blob (*Arrow cursor*) and then decide to accept, reject or copy it by simply clicking the

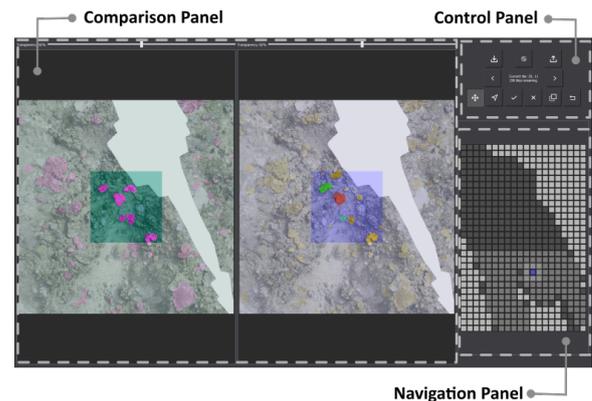


Figure 3: Tool interface. The two windows Comparison Panel (on the left), the Navigation Panel (bottom right) and the Control Panel (top).

dedicated icons. The Comparison Panel shown also the tiles adjacent to the working one to add more context to the user; coral portions on the image edges can be labeled incorrectly. Every time the expert save the changes, the tile will assume a darker color in the Navigation Panel.

If used to validate the semantic segmentation on the *training dataset*, this tool combines the results of the network with the user corrections. So, if a blob is rejected, the corresponding original label is used. False positive in annotation are discarded into as well as false negative predictions are included into the new dataset. Positive predictions better follow the complex contour of the specimens with respect to the original polygonal-based annotation increasing the labeling accuracy.

The proposed tool can be easily generalized with minor modifications to other applications that involve biological/natural structure and require high accurate segmentation results. For example, cells segmentation can take the advantage of the Yes/No paradigm together with the same click-based validation operations for the fast validation of the results. When working on images, the Navigation Panel should be slightly modified to shows each image inside the collection instead of the tiles of the ortho-map.

4. Results

4.1. Annotation time

According to [[PUKF17]] 35.5 seconds are needed to label an object by drawing a bounding box; polygon labeling requires also more time. By using the proposed tool the operator spent approximately 9 hours to compare the result of the segmentation produced by the SegNet (3,530 blobs) with the original annotation (about 1,500 blobs). The user interface demonstrate to be very fast: the time spent is about the 30% of the one of the validation done into a GIS. The average click time is about 9.2s, one third to draw an accurate bounding box. In addition, the information derived by the label comparison is automatically added to new the dataset.

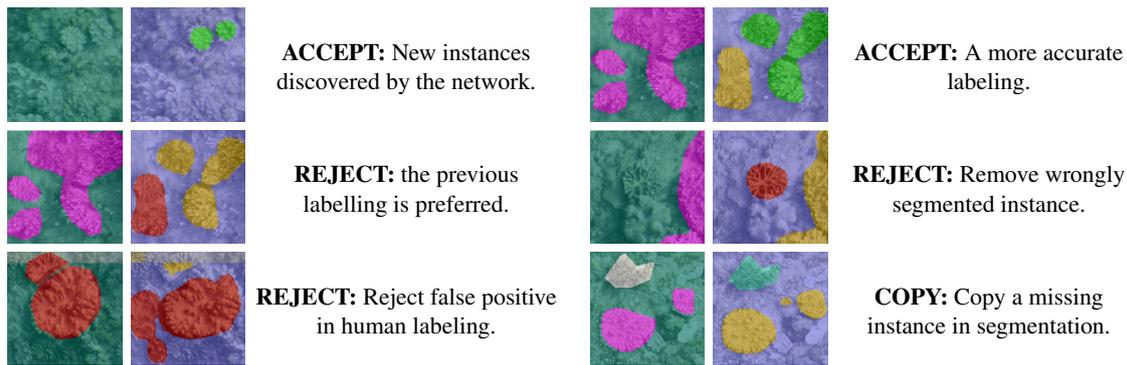


Figure 4: The validation tool operations.

4.2. New dataset and re-training

The new dataset coming from the integrated annotation is characterized by the following:

- Thanks to the new discovered instances, validated as good ones, we gained 4,000,000 pixels of SCD corals, which were false negative (FN) in the original annotation. The increase of the amount of SCD corals w.r.t the original annotation is of about the 16.6%. The higher number of pixels could help a lot the re-training of the network since this dataset is very unbalanced.
- About 1,4 millions of pixels copied by the operators. These quantity represent pixels of missing SCD coral or pixels of incomplete SCD corals (false negative) predicted by the network.
- About the 23% of the pixels marked as false positive predictions of the network; this result (together with missing/incomplete ones) is aligned with network performance, the CNN detects approximately the 70% of the SCD coral present in the test area.
- Only 316,785 pixels marked as false positive predictions of the network and simultaneously false positive in the original annotation (FP generated by the human labeling).

This new dataset have been used to re-train the Bayesian SegNet with the same parameters. We used the Adam optimizer with a fixed learning rate of 10^{-5} with a $L2$ regularization weights of 0.0005. We run the network for 120 epochs. The accuracy increases from 95.1% to 98.3%.

5. Conclusions and Future Work

In this work we propose a simple and effective tool to improve an existing polygonal-based human annotation of marine organisms by integrating the results coming from a semantic segmentation CNN. The time required to improve the dataset is about 1/3 than standard workflow, the new dataset has been successfully integrated with new instances, and the measure of the performance of the network benefit of a more accurate labeling. A Future extension of this tool will be enriched with simple editing operations to improve labeling without compromising the current speed (e.g. a one-click hole filling operation to resolve incomplete segmentation), and with an interface that fully integrates learning algorithms, to obtain a complete semi-automatic annotation tools for the semantic segmentation of natural structures.

References

- [ALKF18] ACUNA D., LING H., KAR A., FIDLER S.: Efficient interactive annotation of segmentation datasets with polygon-rnn++. In *CVPR* (2018). 2
- [BEK*12] BEIJBOM O., EDMUNDS P. J., KLINE D. I., MITCHELL B. G., KRIEGMAN D.: Automated annotation of coral reef survey images. In *CVPR* (2012), pp. 1170–1177. 1
- [BER*15] BEIJBOM O., EDMUNDS P. J., ROELFSEMA C., SMITH J., KLINE D. I., NEAL B. P., DUNLAP M. J., MORIARTY V., FAN T.-Y., TAN C.-J., CHAN S., TREIBITZ T., GAMST A., MITCHELL B. G., KRIEGMAN D.: Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. *PLOS ONE* 10 (07 2015). 1, 2
- [CKUF17] CASTREJÃSN L., KUNDU K., URTASUN R., FIDLER S.: Annotating object instances with a polygon-rnn. In *CVPR* (July 2017), pp. 4485–4493. doi:10.1109/CVPR.2017.477. 2
- [CZLB16] CUI Y., ZHOU F., LIN Y., BELONGIE S.: Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop. In *CVPR* (2016), pp. 1153–1162. 1
- [KBC15] KENDALL A., BADRINARAYANAN V., CIPOLLA R.: Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *CoRR abs/1511.02680* (2015). URL: <http://arxiv.org/abs/1511.02680>, arXiv:1511.02680. 2
- [KBH18] KING A., BHANDARKAR S. M., HOPKINSON B. M.: A comparison of deep learning methods for semantic segmentation of coral reef survey images. In *CVPR 2018 Workshops* (June 2018). 1, 2
- [NDOS03] NINIO R., DELEAN S., OSBORNE K., SWEATMAN H.: Estimating cover of benthic organisms from underwater video images: Variability associated with multiple observers. *Marine Ecology-progress Series - MAR ECOL-PROGR SER* 265 (12 2003), 107–116. 1
- [PCPC17] PALMA M. A., CASADO M. R., PANTALEO U., CERRANO C.: High resolution orthomosaics of african coral reefs: A tool for wide-scale benthic monitoring. *Remote Sensing* 9 (2017), 705. 2
- [PUKF16] PAPADOPOULOS D. P., UIJLINGS J. R. R., KELLER F., FERRARI V.: We don't need no bounding-boxes: Training object class detectors using only human verification. In *CVPR 2016* (2016), pp. 854–863. 1
- [PUKF17] PAPADOPOULOS D. P., UIJLINGS J. R. R., KELLER F., FERRARI V.: Extreme clicking for efficient object annotation. In *ICCV 2017* (2017), pp. 4940–4949. 1, 3