Investigation of hidden parameters influencing the automated object detection in images from the deep seafloor of the HAUSGARTEN observatory

Timm Schoening (tschoeni@cebitec.uni-bielefeld.de), Biodata Mining Group, Faculty of Technology, Bielefeld University
Melanie Bergmann, HGF-MPG Group for Deep-Sea Ecology and Technology,
Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
Antje Boetius, Max Planck Institute for Marine Microbiology, Bremen, Germany
Tim W. Nattkemper, Biodata Mining Group, Faculty of Technology, Bielefeld University

Abstract:
Detecting objects in underwater image sequences and video frames automatically requires the application of selected algorithms in consecutive steps. Most of these algorithms are controlled by a set of parameters, which need to be calibrated for an optimal detection result. Those parameters determine the effectivity and efficiency of an algorithm and their impact is usually well known. There are however further non-algorithmic impact factors (or hidden parameters), which bias the training of a machine learning system as well as the subsequent detection process and thus need to be well understood and taken into account.

In benthic imaging, one dominant, hidden parameter is the distance of the image acquisition device above the seafloor. Variations in the distance lead to variations in the benthic area size being captured, the relative size and position of an object within an image, the effect of the artificial light source and thus the recorded color spectrum. Image processing techniques that allow modeling the induced variations can be used to compensate for those effects and thus allow the exploration of initially biased data. Those processing techniques again require algorithmic parameters, which are influenced by the hidden parameters contained within the initial data.

In supervised machine-learning architectures, further challenges arise from the inclusion of human expert knowledge used for the training of the learning algorithm. Utilizing the knowledge of only one expert can conceal the information needed for the generalization capability of an automated semantic image annotation system. Utilizing the knowledge of several experts requires explicit instruction of the participants to be able to produce comparable results. The fusion of individual expert knowledge poses further hidden parameters that impact the supervised learning architecture. Those could be an individual object specific expertise or the tendency to annotate with more or less self-criticism, which together can be expressed as the expert's trustworthiness.

In the context of mega-fauna detection in benthic images, we investigate the effects of some of these parameters on our machine learning based detection system iSIS [1] that consists of four succeeding steps: Imaging, expert annotation, training, and detection (see Figure 1). The images to be analyzed were taken at the deep-sea, long-term observatory HAUSGARTEN and five experts created an annotation gold standard.

We found, that the hidden parameters from imaging as well as the fusion of expert knowledge could partly be compensated and were able to achieve detection performances of 67% precision and 87% recall. Despite the efforts to compensate the hidden parameters, the detection performance was still varying across the image transect. This poses the potential occurrence of further hidden parameters not taken into account so far.

Here, we correlate the distance of the acquisition device with the image-wise detection results (see Figure 2 A). Also, we show conformity of the automated detection results to the outcome of the manual detection consensus of human experts (see Figure 2 B). Finally, we show the impact of hidden parameters on subsequent steps by means of the effect of image illumination on the human expert annotation.

References:
Schoening, T., Bergmann, M., Ontrup, J., Taylor, J., Dannheim, J., Gutt, J., Purser, A., Nattkemper, T.W. 2012 *Semi-*
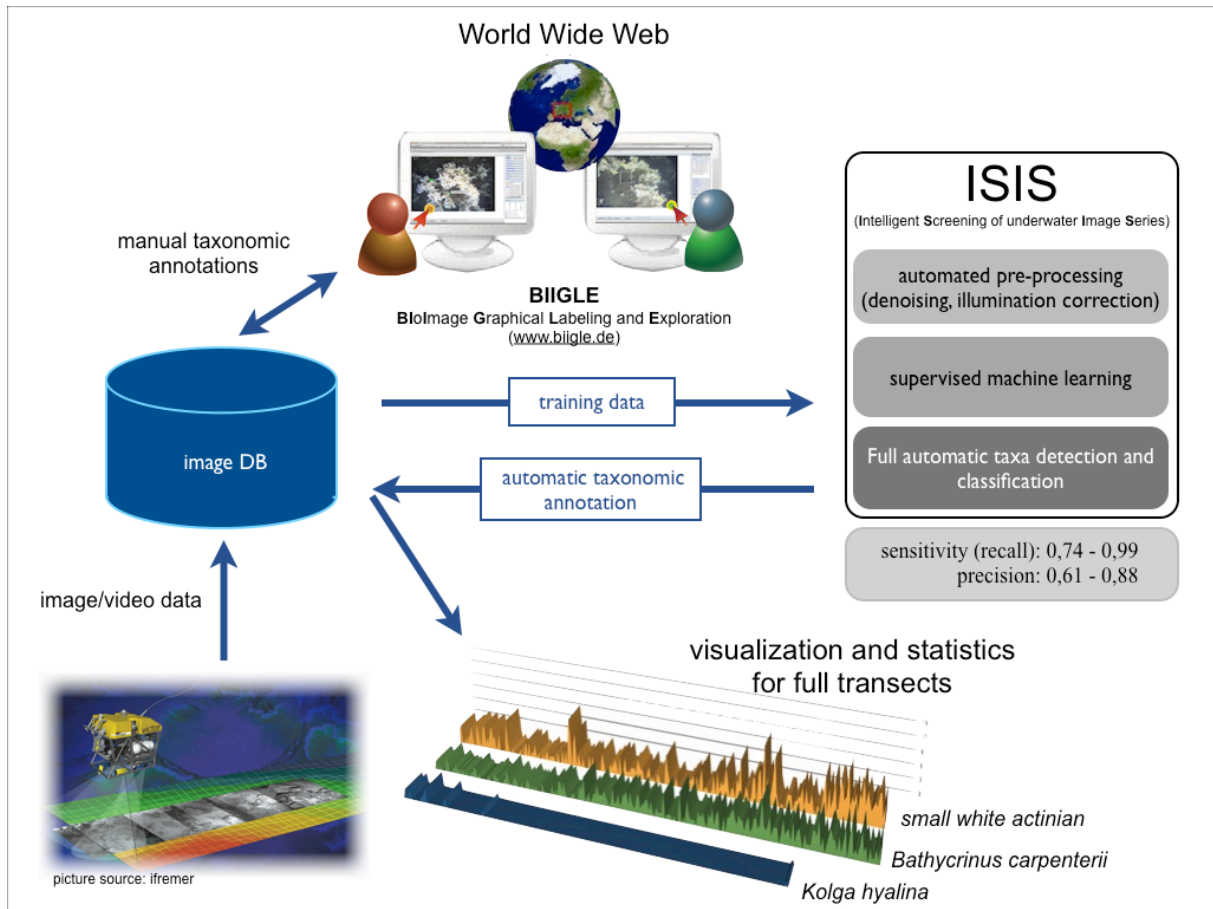
Figure 1: Schematic overview of the automated object detection process. Via the online image annotation software BIIGLE, expert knowledge is gained (top) and stored in a database, together with the images (middle left) that were recorded by an OFOS (bottom left). From those annotations, training data is created and used for the machine-learning step of iSIS (right). The trained SVMs are then used to find further occurrences within previously unseen images from the database. Those occurrence counts can then be visualized over the length of a transect (bottom center).
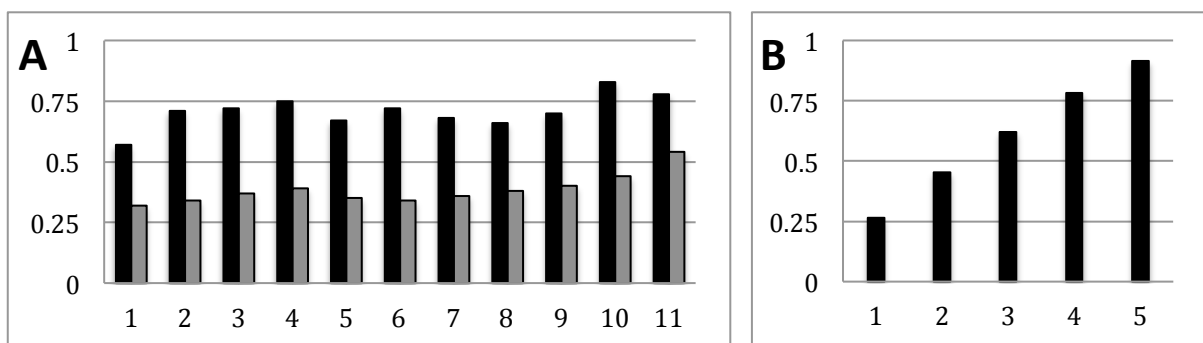
Figure 2: (A) Impact of the image illumination vs. detection performance of all objects in the transect. Each object (detected and annotated) is assigned to one of 11 bins according to its distance to the image`s lightness peak. Bin 1 is closest to the lightness peak, bin 11 the farthest away from it. For all objects within a bin, the average precision (grey bars) and average recall (black bars) are computed. Both values range relatively stable across all bins but nevertheless vary more than 20%.

(B) Impact of the human expert consensus vs. the detection recall. Each object has been annotated by 1 to 5 experts. For those expert-consensus groups, the overall recall is shown. It increases significantly from 0.26 for annotations of a single expert to 0.92 for objects that were annotated by five experts.