



HAL
open science

Automated whistle extraction for precise scaled annotations

Loïc Lehnhoff, Bastien Mérigot, Hervé Glotin

► **To cite this version:**

Loïc Lehnhoff, Bastien Mérigot, Hervé Glotin. Automated whistle extraction for precise scaled annotations. 35th European Cetacean Society Conference, Apr 2024, Catania, Italy. hal-04650230v2

HAL Id: hal-04650230

<https://hal.science/hal-04650230v2>

Submitted on 18 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

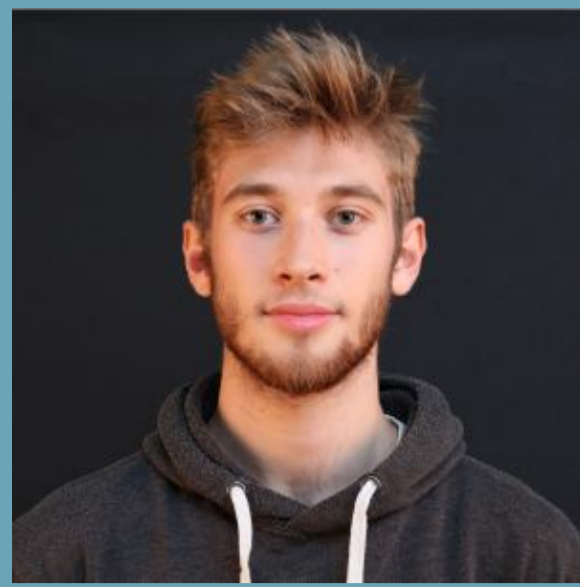
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

AUTOMATED WHISTLE EXTRACTION FOR PRECISE SCALED ANNOTATIONS

AC-10



Loïc Lehnhoff^{a,b,*}, Bastien Mérigot^a, Hervé Glotin^b

^aMARBEC, University of Montpellier, CNRS, IFREMER, IRD, Sete, France

^bUniversity of Toulon, Aix Marseille Univ, CNRS, LIS, CIAN, Toulon, France

*loic.lehnhoff@gmail.com

CONTEXT

- To identify and classify whistles from records is a **time-consuming** and **labor-intensive** process
- Available methods for the automated extraction of whistles contours are prone **to errors**
- Deep learning models require **large quantities of high-quality data** to be trained

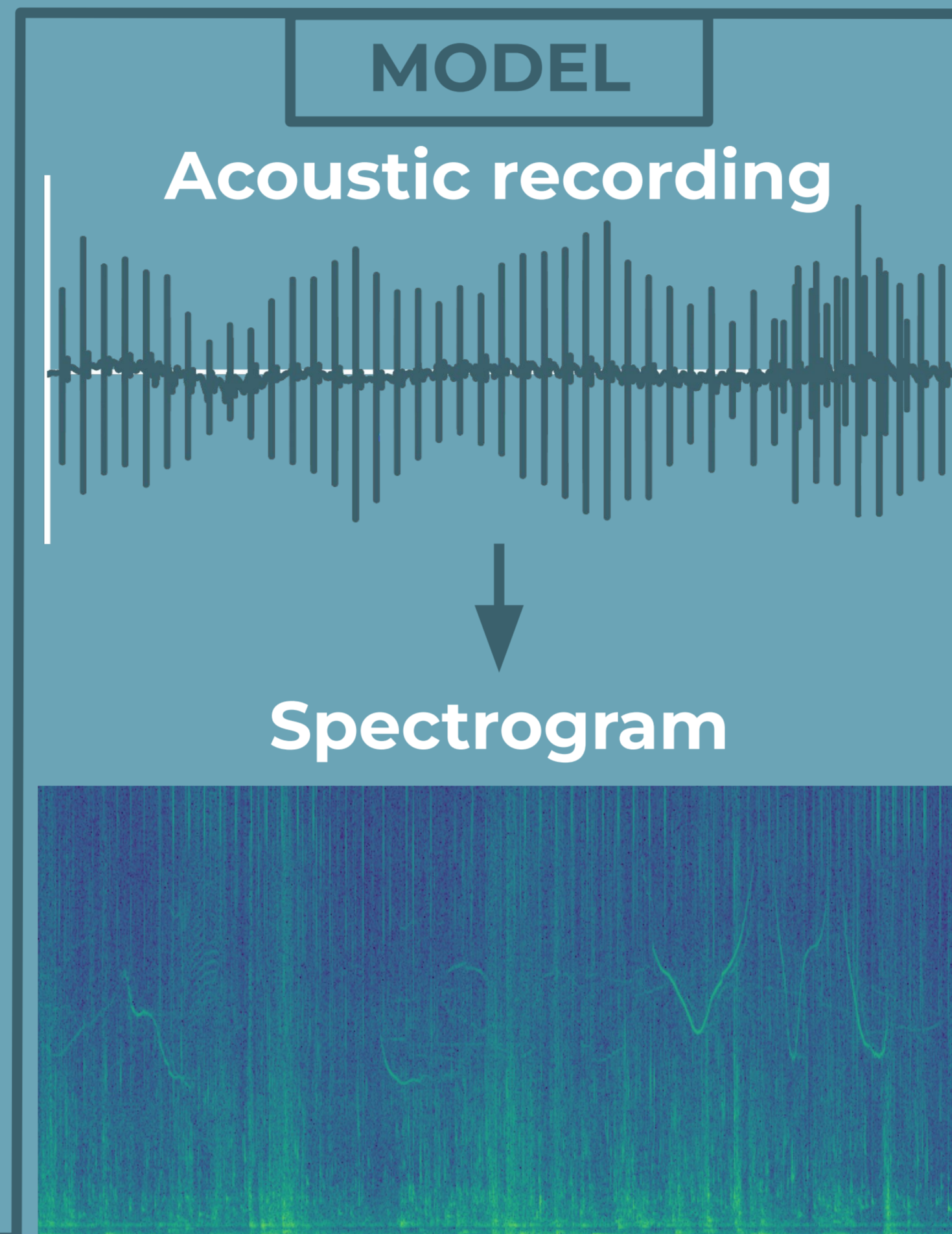
MATERIAL

- 43 minutes of recording (~5% of our data to annotate)
- **1547 whistles contours annotated**
→ using a custom-made annotation tool¹

Acoustic data collected in **Brittany** from experiments with wild short-beaked common dolphins (*Delphinus delphis*)



SCAN ME FOR PDF!



METHODS & RESULTS

YOLOv8^{2,3} is the last state-of-the-art model for **object detection**: it predicts **bounding boxes (bbox)** around the positions of objects in an image.

YOLOv8 is pretrained on **COCO** dataset : >200k real-life images. To use it on spectrograms, we use **transfer learning** to train it to recognise whistles only.

Precision	Recall	mAP50	mAP50-95
68 %	69 %	69 %	42 %

Table 1. 8-fold test scores of YOLOv8 on whistle bbox prediction

Overall satisfying scores :
- Good predictions on isolated whistles
- Difficulties with overlapping whistles

Our **CNN** is simply a pretrained **ResNet18**⁴ (used for image classification) whose last layer has been replaced by a dense layer of neurons in order to predict the coordinates of contours within each bbox, instead of a category.

RMSE in % of the bounding box size:
Errors are small!

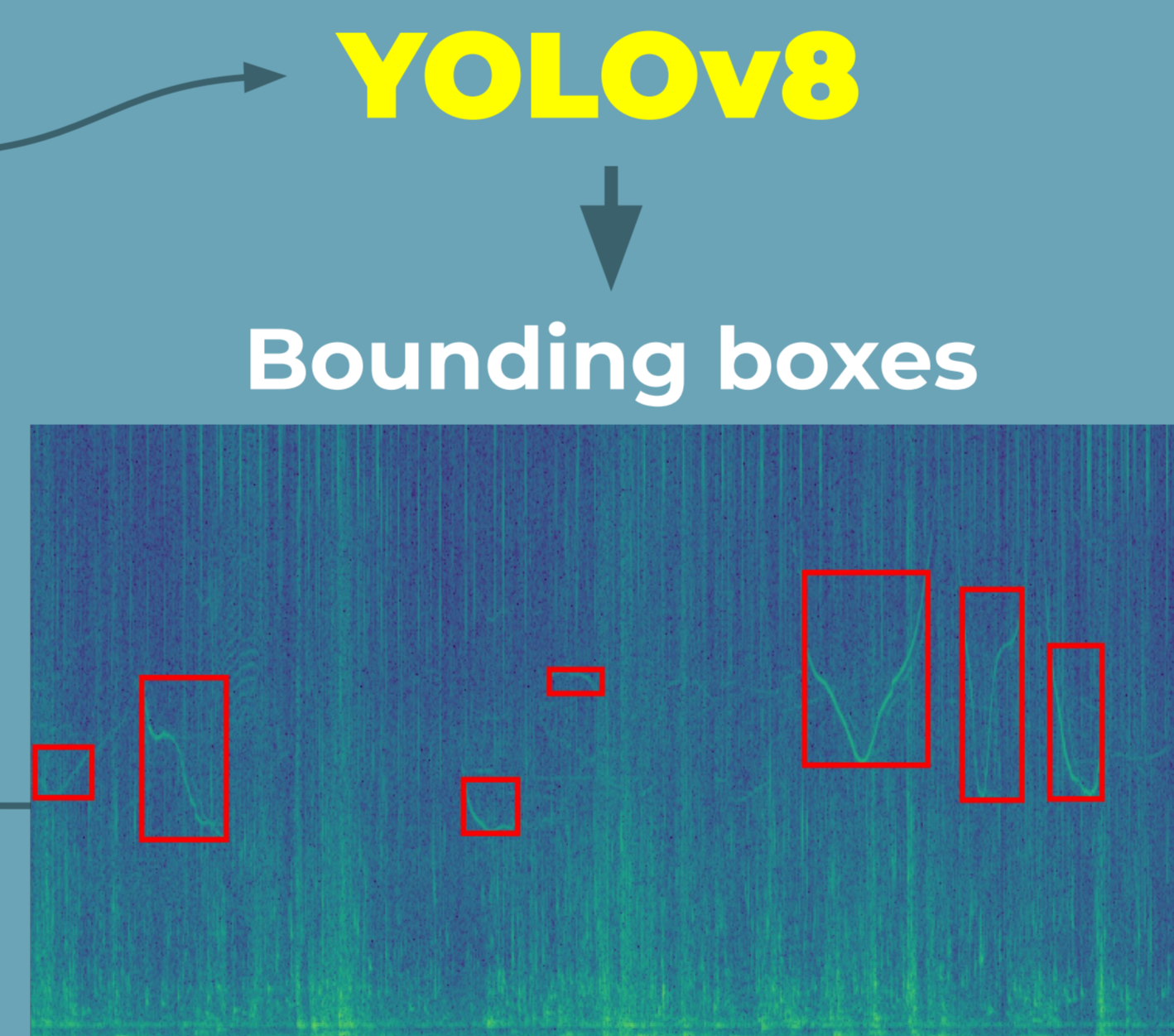
Mean RMSE	Median RMSE
6.1 %	4.6 %

Table 2. 8-fold test scores of the CNN on whistle contour prediction

MANUAL CHECK

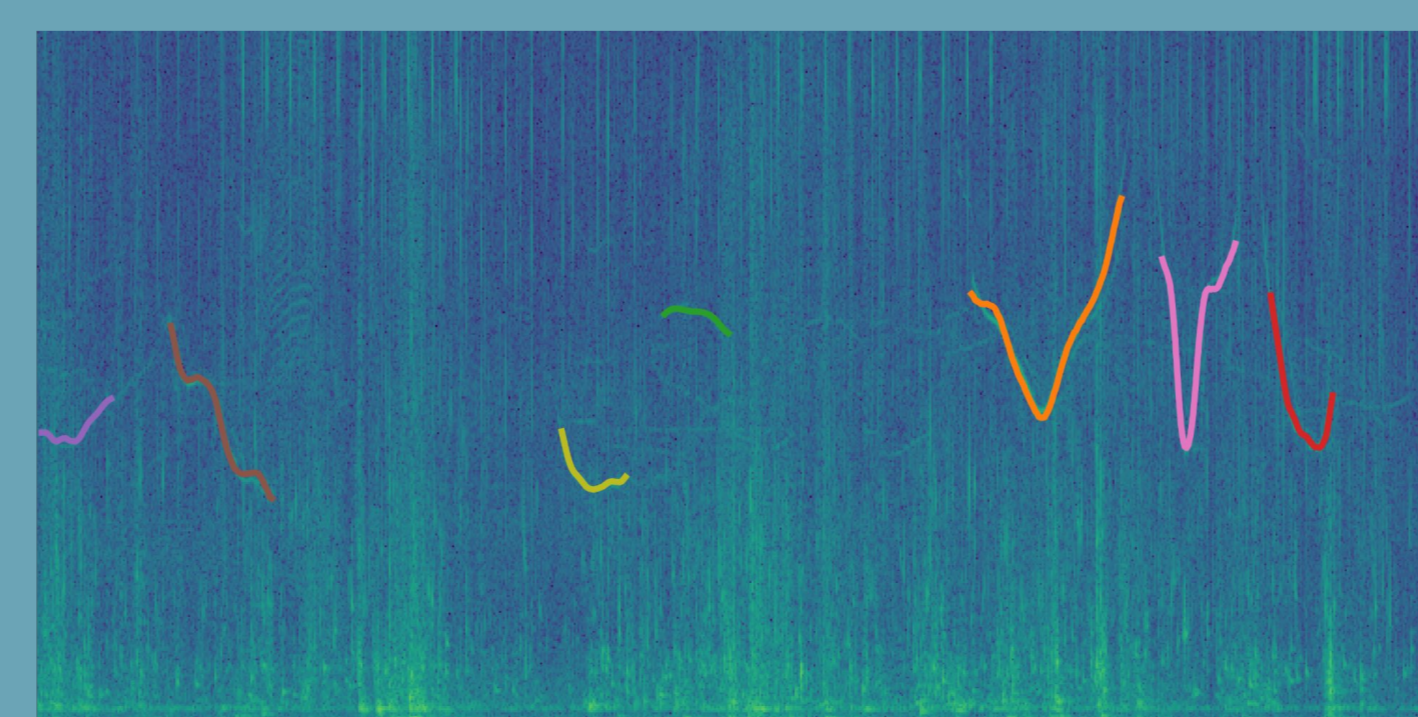
After each prediction step, an **interface** can be displayed to allow for the modification of the results

New bboxes ← **Checking bboxes**

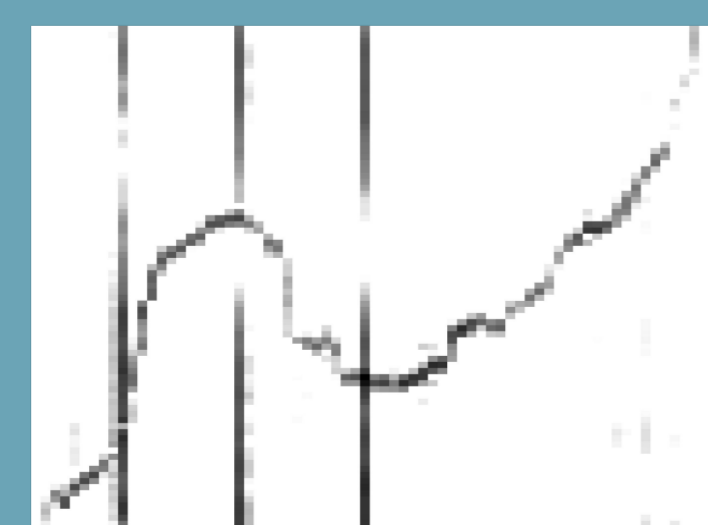
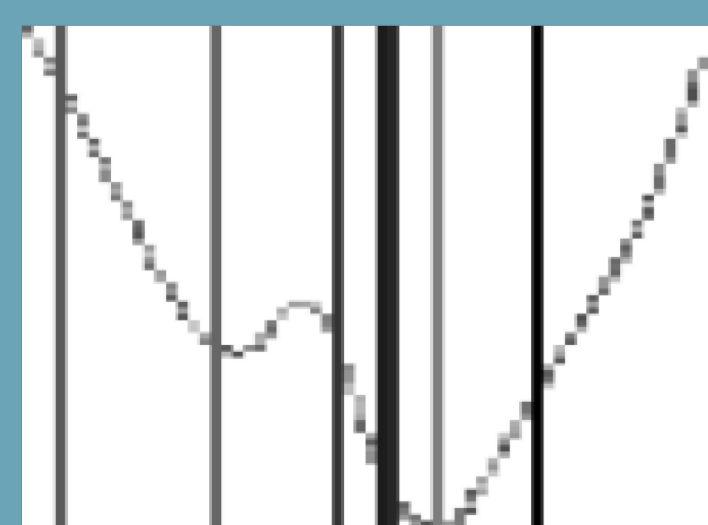


CNN regression

Extracted contours



Our contours had limited variations. To enhance the model's ability to generalize, we added randomly generated images of whistles to the training dataset.



DISCUSSION

Using a **small portion** of our dataset and pretrained AIs, our model is capable of predicting bounding boxes and contours for **isolated whistles with accuracy**.

However, and as we anticipated, when there is **too much noise**, or when whistles are overlapping, **predictions are not satisfying** enough for further use. To overcome this issue, an interface was implemented inside the prediction process in order to make fast and precise modifications to any errors the model could make.

The model could be further improved with the use of a greater number of overlapping whistles during training, which are the most difficult to extract with accuracy, even for a human annotator.

Ack: H.G. thanks ANR for grants ULPCochlea ANR-21-CE04-0020 & DGA IAD for AI Chair ADSIL ANR-20-CHIA-0014. B.M. thanks EMFF and FFP for their fundings. L.L. thanks the University of Montpellier for its PhD's grant.

1. Lehnhoff, L. PyAVA [computer software]. 2024. <https://gitlab.lis-lab.fr/loic.lehnhoff/PyAVA>
 2. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. "You only look once: Unified, real-time object detection." Proceedings of the IEEE CVPR conference. 2016. <https://arxiv.org/abs/1506.02640>
 3. Jocher, G., Chaurasia, A. & Qiu, J. Ultralytics YOLO version 8.0.0 [computer software]. 2023. <https://github.com/ultralytics/ultralytics>
 4. He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE CVPR conference. 2015. <https://arxiv.org/abs/1512.03385>
 5. Chavin S., Mahé P., Hermet T., Deloustal N. & Glotin H. Analyse automatisée de la diversité acoustique, de la détection d'espèces aux indices bioacoustiques. RR LIS CNRS. 2023. http://sabiod.lis-lab.fr/pub/LIS_QUEBEC_RR.pdf

