A Human-in-the-Loop Tool for Annotating Passive Acoustic Monitoring Datasets (Extended Abstract)

Hannes Kath[1,2], Thiago S. Gouvêa[1], and Daniel Sonntag[1,2]

[1] German Research Centre for Artificial Intelligence (DFKI), Oldenburg, Germany
`{hannes.kath, thiago.gouvea, daniel.sonntag}@dfki.de`
[2] University of Oldenburg, Applied Artificial Intelligence (AAI), Oldenburg, Germany

Abstract. Passive Acoustic Monitoring (PAM) has become a key technology in wildlife monitoring, generating large amounts of acoustic data. However, the effective application of machine learning methods for sound event detection in PAM datasets is highly dependent on the availability of annotated data, which requires a labour-intensive effort to generate. This paper summarises two iterative, human-centred approaches that make efficient use of expert annotation time to accelerate understanding of the data: Combining transfer learning and active learning, we present an annotation tool that selects and annotates the most informative samples one at a time [11]. To annotate multiple samples simultaneously, we present a tool that allows annotation in the embedding space of a variational autoencoder manipulated by a classification head [10]. For both approaches, we provide no-code web applications for intuitive use by domain experts.

Keywords: Passive Acoustic Monitoring · Active Learning · Transfer Learning· Deep Generative Model · Annotation Tool

1 Introduction

Biodiversity loss is among the most pressing issues of our days [1]. Drivers of the negative change have been accelerating, and meeting internationally agreed conservation targets will require transformative change [7]. While machine learning methods have been increasingly brought to bear to support wildlife management, the tools available to those on the ecological front lines still lag the state-of-the-art in artificial intelligence research [6,21]. Passive acoustic monitoring (PAM) is a powerful technology for wildlife monitoring, allowing ecologists to gather extensive data on wildlife with minimal disturbance of habitats [18,19]. PAM systems can be used to continuously record sounds from various biomes, offering valuable insights into animal behaviour, species richness, and ecosystem health, with applications in ecosystem management and rapid biodiversity assessments [17].

However, effectively utilising this vast amount of data still poses significant management and analysis challenges. Sound event detection in particular (e.g., species identification) is limited by the need for annotated data to train supervised machine learning models. PAM data annotation is usually done manually, in a laborious and time-consuming process: domain experts listen to each audio file, annotating events by manually selecting time segments on a graphical representation of the sound (e.g., amplitude envelope or spectrogram) [2,16,20]. This approach is incompatible with the large volume of data generated by PAM. Seadash introduces a graphical implementation of data programming, but lacks evaluation on real life datasets [5]. Scikit-maad [22] and BamScape [14] are tools for large scale PAM data analysis by spectrogram segmentation and clustering; as command line tools, they lack interactivity. A promising, under-explored direction of research on interactive PAM annotation is the leveraging of small amounts of available labels (e.g., at early stages of annotation of a novel dataset) to improve the efficiency of the annotation process.

In the following, we present the core ideas of two human-centred approaches that efficiently annotate single or multiple samples of PAM datasets by iteratively incorporating previously annotated data.

2 Single sample annotation tool

Active sampling significantly reduces the number of samples required to train a classifier compared to passive sampling. We use a combination of transfer learning and active learning for sample selection. The unlabelled dataset is embedded using BirdNet, a neural model trained on focal recordings of songbirds [8], which has shown superior performance as an embedding model for PAM data [3,12]. A number of uncertainty-based and diversity-based sampling strategies for multi-label active learning are then used to select segments for annotation (see [12] for details). While the user retains the power to choose which samples to annotate (fig. 1, left), the pipeline uses all available labels to recommend the most informative samples. Samples are presented as spectrograms and filterable audio (fig. 1, centre) and annotated with graphical interactions (fig. 1, right). Evaluations by a domain expert on a real life dataset show that using the interface results in a time improvement of a factor of 2-4 compared to conventional tools [11].

3 Multi sample annotation tool

Semantic clustering of unlabelled data allows multiple samples to be annotated simultaneously. Our model architecture is inspired by [9,15] and extends a variational autoencoder (VAE) [13] with a classification head to adjust the latent space to represent events as outliers. The input data X is processed by the encoder and mapped to the 2D-latent variable Z, which is presented to the user as a state-space and represents the input to the decoder and classifier. The decoder computes the reconstruction $\tilde{X}$ for the whole dataset. The classifier processes only annotated data by computing predicted labels $\tilde{Y}$ from Z for all files with a label Y, grouping data points of the same category into Z. The VAE and classifier are jointly optimised by minimising the loss function

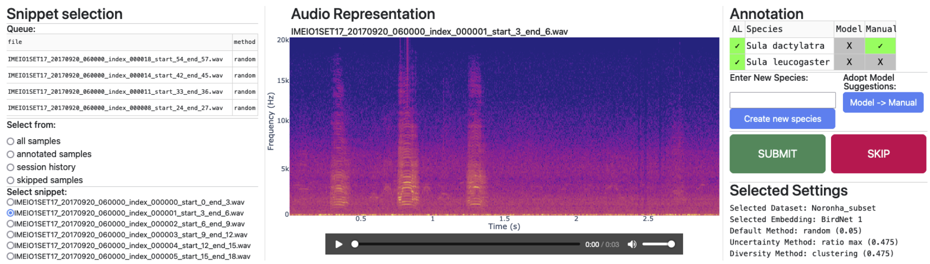

Fig. 1: Single sample annotation tool layout. Left: Sample selection column. Centre: Sample representation as spectrogram and audio. Right: Annotation column.

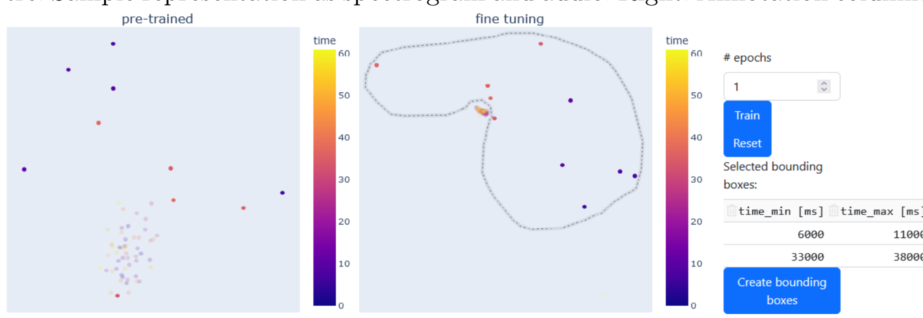

Fig. 2: Multi sample annotation tool layout. Left: Embedding space pre-trained. Centre: Embedding space fine-tuned. Right: Generated annotation table.

$\mathcal{L} = \mathcal{L}_{\text{reconst}}(X, \tilde{X}) + D_{KL}\left(q_\phi(Z \mid X) \mid\mid p(Z)\right) + \mathcal{H}(Y, \tilde{Y})$. The first two terms are as in [13], and $\mathcal{H}$ is the cross entropy between Y and $\tilde{Y}$. Figure 2 (left) shows the latent space of the pure VAE. Adding the classification head (fig. 2, centre) distorts the latent space to represent events as outliers, which can be encircled by the user and further annotated using the generated table (fig. 2, right). Evaluations on a synthetic dataset show that using the fine-tuned space for annotation outperforms the pre-trained model with an F-score of 94.2 % compared to 77.0 %.

4 Conclusion

Our work highlights the need for accessible machine learning-based annotation of PAM datasets and proposes two interactive, human-centred tools. For single sample annotation, we present an annotation tool that uses a combination of transfer learning and active learning. For multi sample annotation, we propose to modify the latent space of a VAE by adding a classifier head that generates an actionable, low-dimensional representation space of the input data.

Preliminary evaluations of the tools show promising results, and future work includes integrating our tool into real-world projects [4].

Acknowledgements

This research is part of the Computational Sustainability & Technology project area[3], and has been supported by the Ministry for Science and Culture of Lower Saxony (MWK), the Endowed Chair of Applied Artificial Intelligence, Oldenburg University, and DFKI.

References

1. Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P.A., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A., Daily, G.C., Loreau, M., Grace, J., Larigauderie, A., Srivastava, D.S., Naeem, S.: Biodiversity loss and its impact on humanity. Nature **486**, 59–67 (2012)
2. Cañas, J., Toro-Gómez, M., Sugai, L., et al.: A dataset for benchmarking neotropical anuran calls identification in passive acoustic monitoring. Scientific Data **10**(1), 771 (2023)
3. Ghani, B., Denton, T., Kahl, S., Klinck, H.: Global birdsong embeddings enable superior transfer learning for bioacoustic classification. Scientific Reports **13**(1), 22876 (2023)
4. Gouvêa, T.S., Kath, H., Troshani, I., Lüers, B., Serafini, P.P., Campos, I., Afonso, A., Leandro, S.M.F.M., Swanepoel, L., Theron, N., Swemmer, A.M., Sonntag, D.: Interactive machine learning solutions for acoustic monitoring of animal wildlife in biosphere reserves. IJCAI 2023 (2023)
5. Gouvêa, T.S., Troshani, I., Herrlich, M., Sonntag, D.: Annotating sound events through interactive design of interpretable features. In: Schlobach, S., Pérez-Ortiz, M., Tielman, M. (eds.) HHAI 2022: Augmenting Human Intellect - Proceedings of the First International Conference on Hybrid Human-Artificial Intelligence, Amsterdam, The Netherlands, 13-17 June 2022. Frontiers in Artificial Intelligence and Applications, vol. 354, pp. 305–306. IOS Press (2022). `https://doi.org/10.3233/FAIA220225`
6. Gouvêa, T., Kath, H., Troshani, I., et al.: Interactive Machine Learning Solutions for Acoustic Monitoring of Animal Wildlife in Biosphere Reserves. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. pp. 6405–6413. Macau, SAR China (2023)
7. IPBES: Summary for policymakers of the global assessment report on biodiversity and ecosystem services. Tech. rep., Zenodo (Nov 2019)
8. Kahl, S., Wood, C., Eibl, M., Klinck, H.: BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics **61**, 101236 (2021). `https://doi.org/10.1016/j.ecoinf.2021.101236`
9. Kath, H., Gouvêa, T.S., Sonntag, D.: A deep generative model for interactive data annotation through direct manipulation in latent space. arXiv:2305.15337 (2023)
10. Kath, H., Gouvêa, T., Sonntag, D.: A Human-in-the-Loop Tool for Annotating Passive Acoustic Monitoring Datasets. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. pp. 7140–7144. Macau, SAR China (2023). `https://doi.org/10.24963/ijcai.2023/835`

[3] `https://cst.dfki.de/`

11. Kath, H., Serafini, P.P., Campos, I.B., Gouvêa, T.S., Sonntag, D.: Demo: Enhancing wildlife acoustic data annotation efficiency through transfer and active learning. In: Proceedings of the 33nd International Joint Conference on Artificial Intelligence, IJCAI (2024)
12. Kath, H., Serafini, P.P., Campos, I.B., Gouvêa, T.S., Sonntag, D.: Leveraging Transfer Learning and Active Learning for Sound Event Detection in Passive Acoustic Monitoring of Wildlife. In: 3rd Annual AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE). Vancouver, Canada (2024)
13. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: Bengio, Y., LeCun, Y. (eds.) 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings (2014)
14. Michaud, F., Sueur, J., Cesne, M.L., Haupert, S.: Unsupervised classification to improve the quality of a bird song recording dataset. ArXiv **abs/2302.07560** (2022)
15. Paige, B., Siddharth, N., van de Meent, J., Desmaison, A., Goodman, N.D., Kohli, P., Wood, F.D., Torr, P.H.S.: Learning disentangled representations with semi-supervised deep generative models. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA. pp. 5925–5935 (2017)
16. Perry, S., Tiwari, V., Balaji, N., Joun, E., Ayers, J., Tobler, M., Ingram, I., Kastner, R., Schurgers, C.: Pyrenote: a web-based, manual annotation tool for passive acoustic monitoring. In: IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems, MASS 2021, Denver, CO, USA, October 4-7, 2021. pp. 633–638. IEEE (2021). `https://doi.org/10.1109/MASS52906.2021.00091`
17. Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S.: Rapid Acoustic Survey for Biodiversity Appraisal. PLOS ONE **3**(12), e4065 (2008). `https://doi.org/10.1371/journal.pone.0004065`
18. Sugai, L., Llusia, D.: Bioacoustic time capsules: Using acoustic monitoring to document biodiversity. Ecological Indicators **99**, 149–152 (2019). `https://doi.org/10.1016/j.ecolind.2018.12.021`
19. Sugai, L., Silva, T., Ribeiro, J., Llusia, D.: Terrestrial Passive Acoustic Monitoring: Review and Perspectives. BioScience **69**(1), 15–25 (2019). `https://doi.org/10.1093/biosci/biy147`
20. Tkachenko, M., Malyuk, M., Holmanyuk, A., Liubimov, N.: Label Studio: Data labeling software (2020)
21. Tuia, D., Kellenberger, B., Beery, S., Costelloe, B.R., Zuffi, S., Risse, B., Mathis, A., Mathis, M.W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I.D., van Horn, G., Crofoot, M.C., Stewart, C.V., Berger-Wolf, T.: Perspectives in machine learning for wildlife conservation. Nature Communications **13**(1), 792 (Feb 2022), number: 1 Publisher: Nature Publishing Group
22. Ulloa, J.S., Haupert, S., Latorre, J.F., Aubin, T., Sueur, J.: scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python. Methods in Ecology and Evolution pp. 2041–210X.13711 (Sep 2021). `https://doi.org/10.1111/2041-210X.13711`