Towards Domain Shift in Location-Mismatch Scenarios for Bird Activity Detection

Fraunhofer IDMT

Amir Latifi Bidarouni, Jakob Abeßer

Semantic Music Technologies Group, Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany

Introduction

- Bioacoustics monitoring provides insight of wildlife wellbeing
- Deep learning methods help analyzing large amount of data

Challenges:

- Performance drop of trained model because of domain shift between **source domain** (**SD**) and **target domains** (**TD**'s)
  - **SD**: Dataset used to train the model
  - **TD**: Dataset upon which the trained model will be used to perform predictions
  - **Domain shift**: Differences in acoustic data distribution between SD and TDs due to various recording conditions such as varying location or microphone

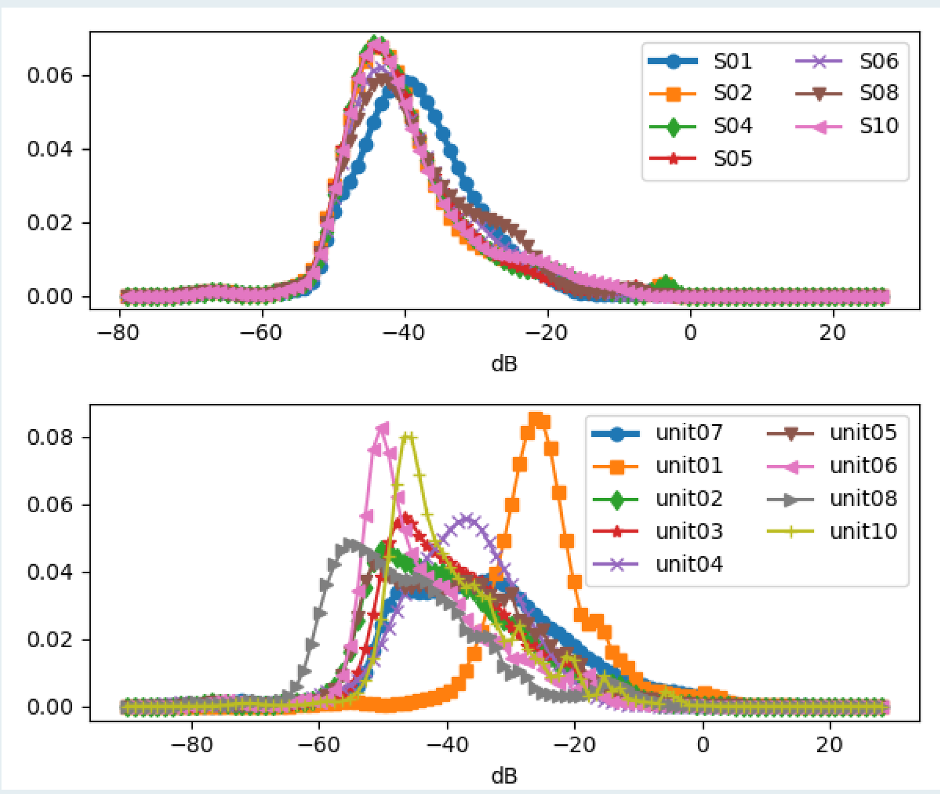
Research Objectives:

- **Comparison of domain adaptation (DA) techniques** for bird activity detection (binary classification)
  - DA: methods to reduce domain shift
- **Domain shift measurement** using various distance and divergence metrices on:
  - Overall dynamic range
  - Energy distribution across frequency bands

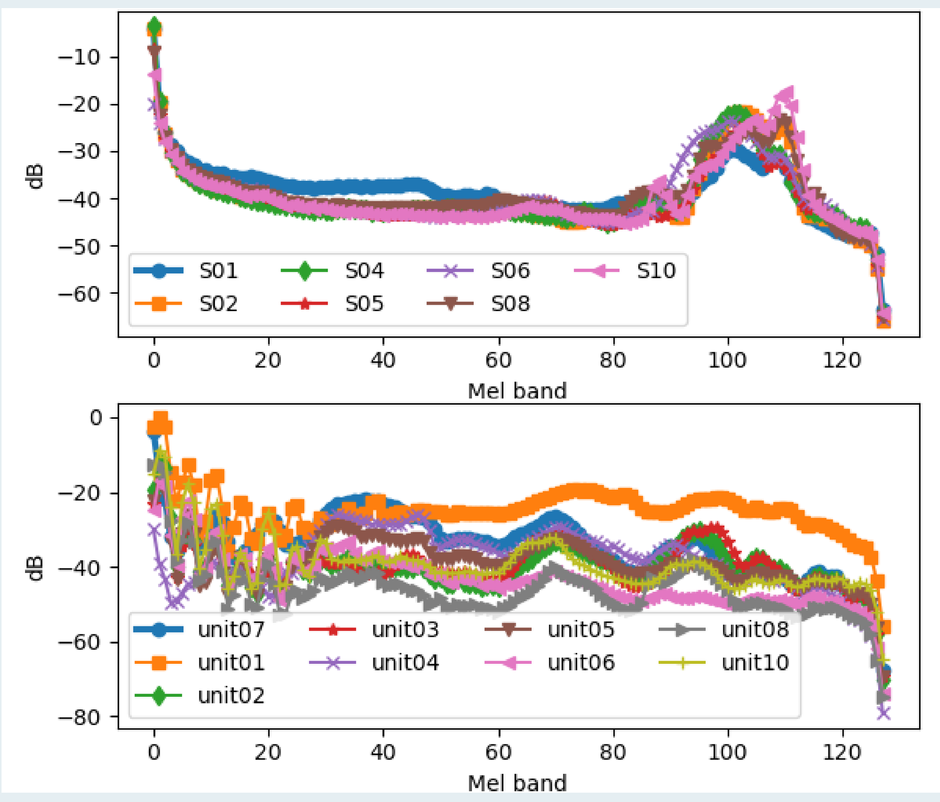
Datasets:

- BirdVox-296h (148 2h audio recordings, 9 locations)
- Southwestern Amazon Basin (21h, 7 locations)

Probability distribution of raw data for different recording locations for AMZ (top) and BV (bottom).



Energy distribution across frequency bands for different recording locations for AMZ and BV.



Methodology

Feature Extraction:

- 44.1 kHz sample rate, 128 Mel bands, FFT size = 2048, window size = 1024, hop-size = 512, patch-size = 10s

Model Architecture:

- CNN420 [1], batch size=32, Adam optimizer (lr=0.001)

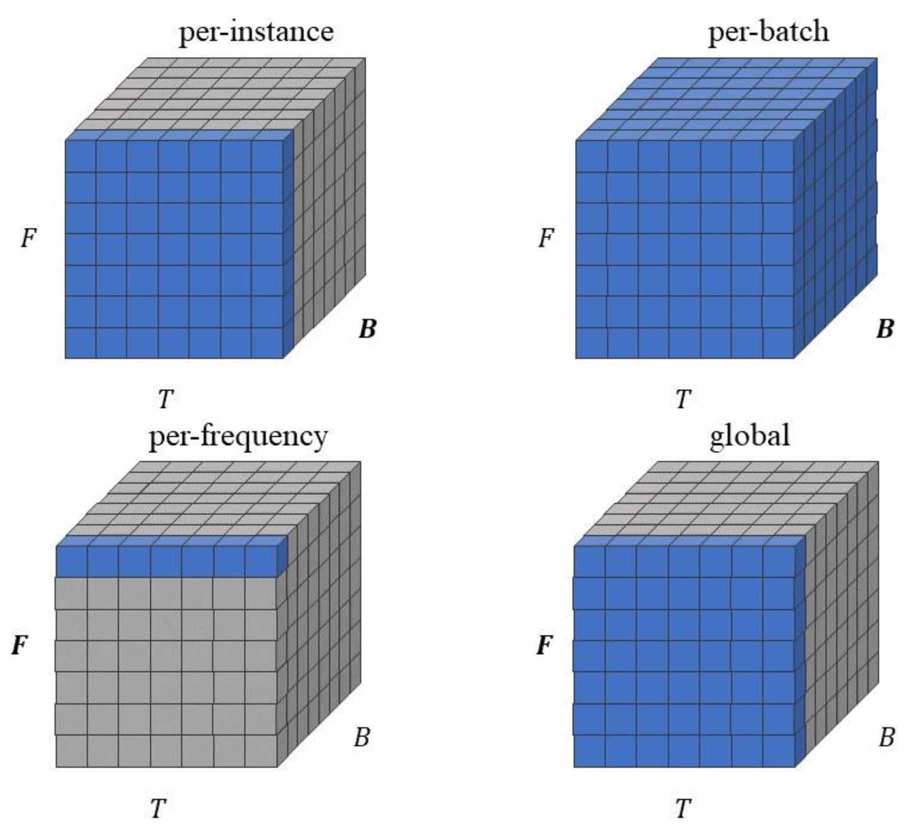
Domain adaptation methods:

- Z-score normalization (standardization)
- Relaxed Instance Frequency Normalization
- Instance-Wise Feature Projection-Based Domain Adaptation[2]

Distance and divergence measurements methods:

- KL-divergence (KLD)
- Wasserstein distance (WSD)
- Cramér-Von Mises distance (CVM)

Data Partitioning [2]:



- Within / Between domains

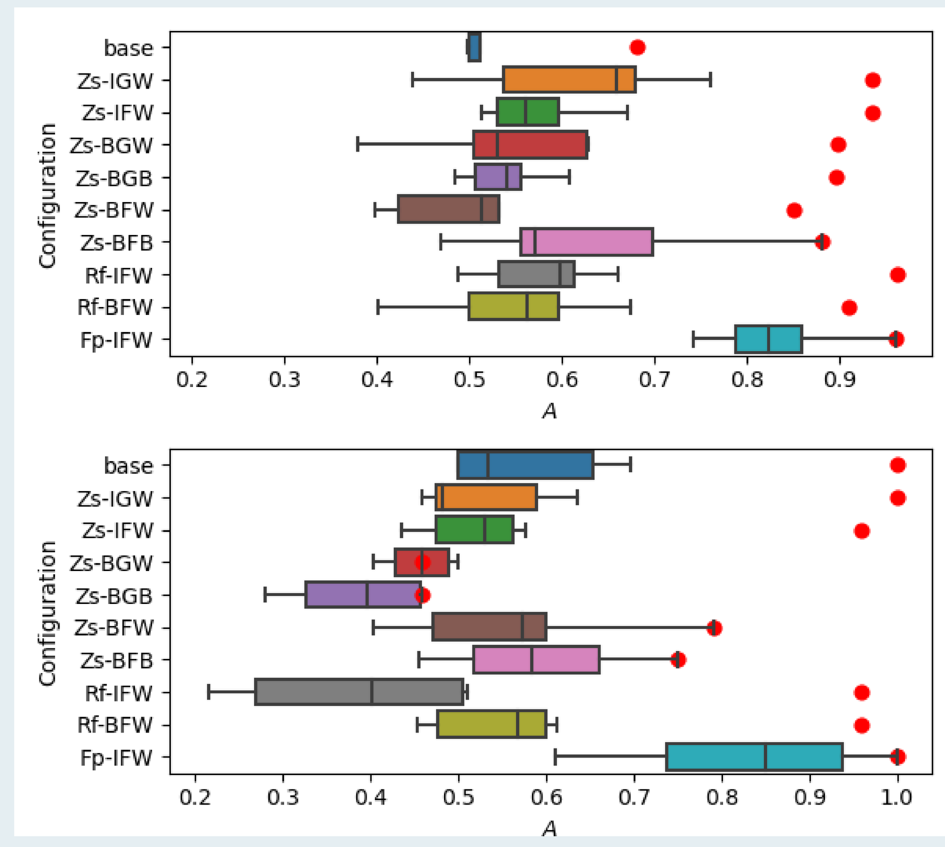
Evaluation

Experiment1 – Comparison of DA techniques:

Configurations with the corresponding normalization methods and dataset partition approach.

Configuration	Normalization Method	Partitioning Dimension: Batch *(B)*	Frequency *(F)*	Domain
Base	-	-	-	-
Zs-IGW	Z-score	instance	global	within
Zs-IFW	Z-score	instance	frequncy	within
Zs-BGW	Z-score	batch	global	within
Zs-BGB	Z-score	batch	global	between
Zs-BFW	Z-score	batch	frequncy	within
Zs-BFB	Z-score	batch	frequncy	between
Rf-IFW	RFN	instance	frequncy	within
Rf-BFW	RFN	batch	frequncy	within
Fp-IFW	FPDA (IFPDA)	instance	frequncy	within

Accuracy for AMZ (top) and BV (bottom) datasets.



- Some configurations showing lower *A* than base on SD or TD due to increased domain shift or poor generalization
- IFPDA shows superior performance on both SD and TD

Experiment 2 – Quantifying Domain shift:

Huber regression coefficients (HRC) and coefficient of determination R2 values between accuracy drop and domain shift measurement metric.

Metric	Data Representation	BV dataset HRC	BV dataset R2	AMZ dataset HRC	AMZ dataset R2
MSE	Magnitude	260580	-0.41	446414	-0.68
MSE	Frequency	42.59	-0.11	94.37	**0.10**
KLD	Magnitude	147.16	**0.08**	192.83	**0.01**
WSD	Magnitude	5365.7	-0.31	8533	-0.73
WSD	Frequency	82.36	**0.21**	61.75	-0.25
CVM	Magnitude	22.48	-2.71	15.72	-1.66
CVM	Frequency	65.77	-1.86	3390	**0.36**

Conclusions

- The IFPDA method is the most efficient of the investigated DA methods for the two investigated bioacoustics datasets
- Energy distribution across frequency bands within a dataset provides a better representation for domain shift measurement compared to global dynamic range
- In general, positive regression coefficients confirm our intuition that the accuracy drop from the SD to the TD increases with increasing domain shift

References

- [1] D. Johnson and S. Grollmisch, "Techniques improving the robustness of deep learning models for Industrial Sound Analysis," in Proceedings of the 2020 European Signal Processing Conference (EUSIPCO), Online, 2021, pp. 81–85
- [2] A. L. Bidarouni and J. Abeßer, "Unsupervised feature-space domain adaptation applied for audio classification," in 2023 4th International Symposium on the Internet of Sounds, Pisa, Italy, 2023, pp. 1–7

Bio Monitor 4CAP

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation