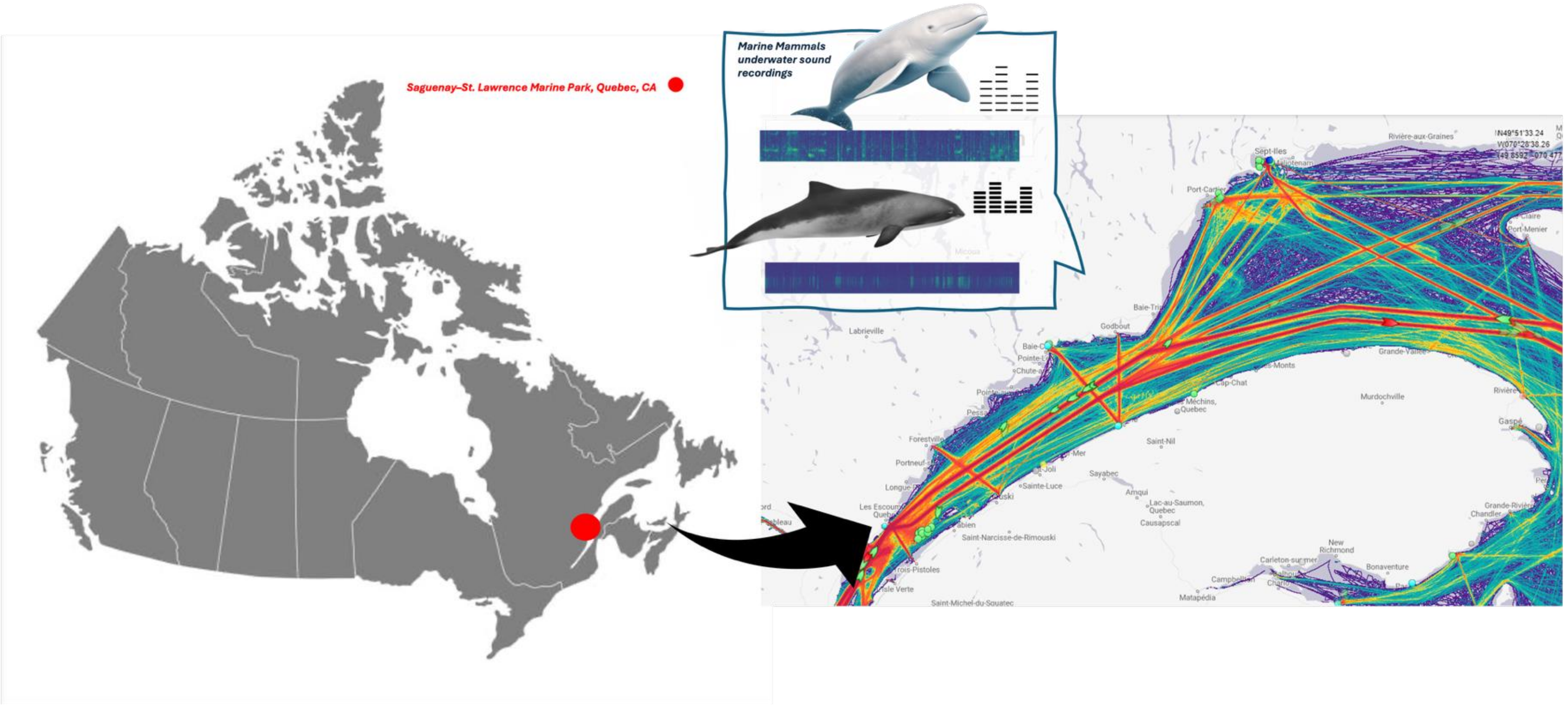
Bioacoustic Multi-Step Attention : Underwater Ecosystem Monitoring in Climate Change Context

Amine Razig[1,5], Youssef Soulaymani[2], Loubna Benabbou[3], Pierre Cauchy[4]
[1] MILA – Quebec AI Institute, [2] Université de Montréal, [3] Université du Québec à Rimouski, [4] ISMER, [5] Institut Polytechnique de Paris

General context



St. Lawrence Estuary: a unique and ecologically sensitive region. One of the world's largest navigable waterways.

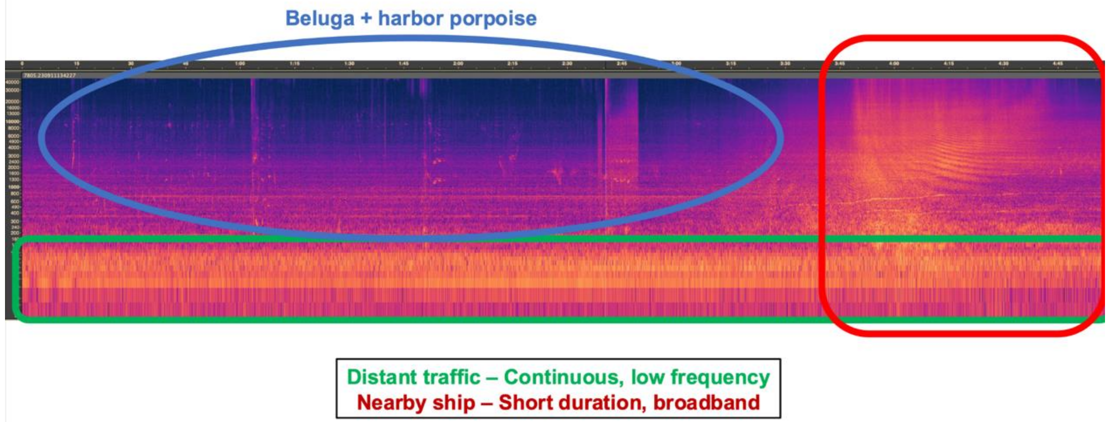
Increasing pressures:

- Shipping traffic ⇒ acoustic pollution
- Climate change ⇒ altered soundscapes and species distribution

Problem Statement and Objectives

In addition to being a high-traffic area, St. Lawrence Estuary is critical acoustic habitat for marine mammals

- Main challenges: Noisy conditions and overlapping signals



- General objective:

  Automatically detect and classify marine mammal vocalizations in noisy, multi-species recordings (Porpoise and Beluga Clicks, Whistles)

Problem Statement and Objectives

The available data comes from hydrophone networks that continuously record underwater soundscapes.

Dataset: Exclusive subset of the SSLMP monitoring program

- ~1,500 h passive acoustic monitoring (hydrophones)
- ~ 500 h shore-based visual surveys
- ~10,000 annotated 5-min segments (Beluga, Harbour Porpoise)
- Frequency range: 10 Hz – 150 kHz (whistles, clicks, vessel noise)

- **multi-label setup** (Whistle, Beluga Click, Porpoise Click)

→

- **multi-class** by encoding all label combinations ($2^3 = 8$ categories)

Problem Statement and Objectives

- In order to study mammals, we must first establish **a reliable methodology for detecting them.**
- End-to-end multi-step framework combining spectrograms + pseudo attention masks

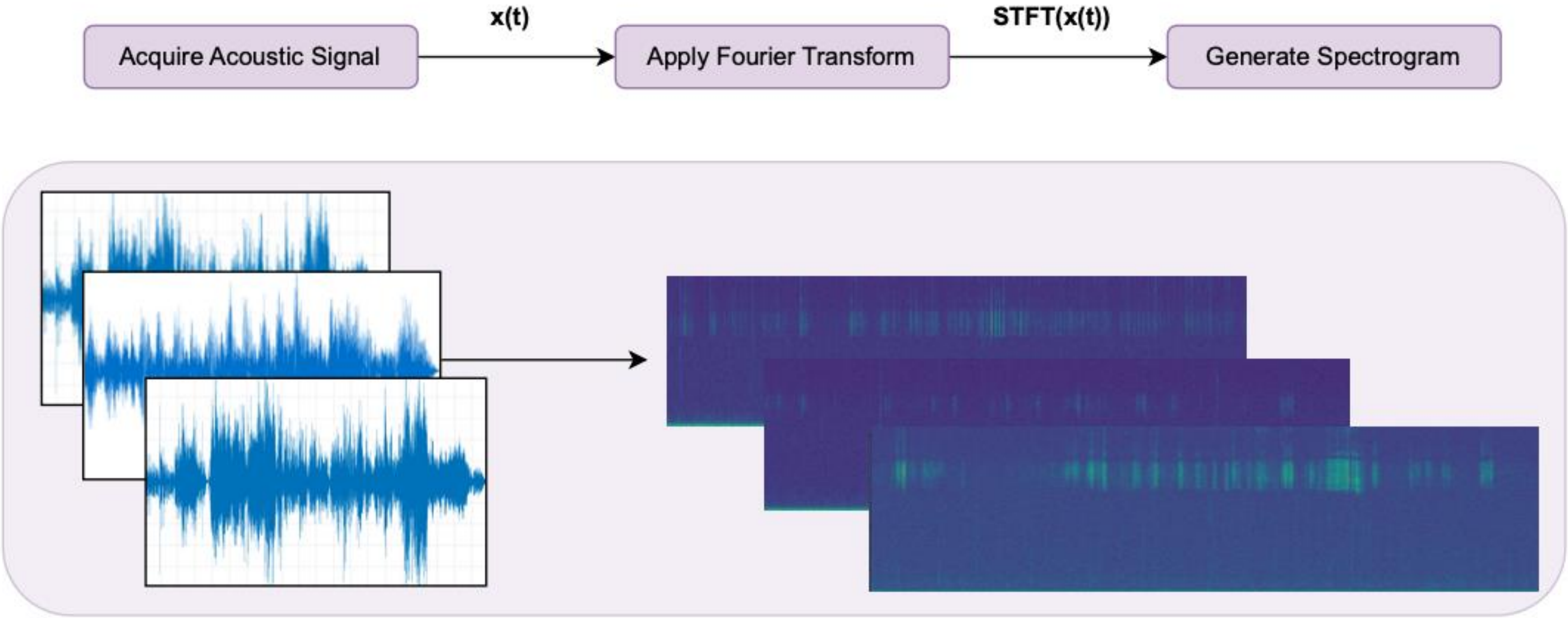
Problem Setup:

$$\hat{y} = \mathcal{C}\Big(\mathrm{Fuse}\Big(\mathcal{E}_{\mathrm{spec}}(\mathcal{T}(x(t))),\ \mathcal{E}_{\mathrm{mask}}(\mathcal{M}_{\mathrm{seg}}(\mathcal{T}(x(t))))\Big)\Big), \quad \hat{y} \in \mathbb{R}^K$$

where $\mathcal{T}$ is the STFT , $\mathcal{E}_{\mathrm{spec}}$ and $\mathcal{E}_{\mathrm{mask}}$ are the embedding functions for the spectrogram and mask, respectively, and $\mathrm{Fuse}(\cdot,\cdot)$ denotes the mid-level embedding fusion. $\mathcal{M}_{\mathrm{seg}}$ a segmentation model and $\mathcal{C}$ a classifier. ▸ see

Framework description

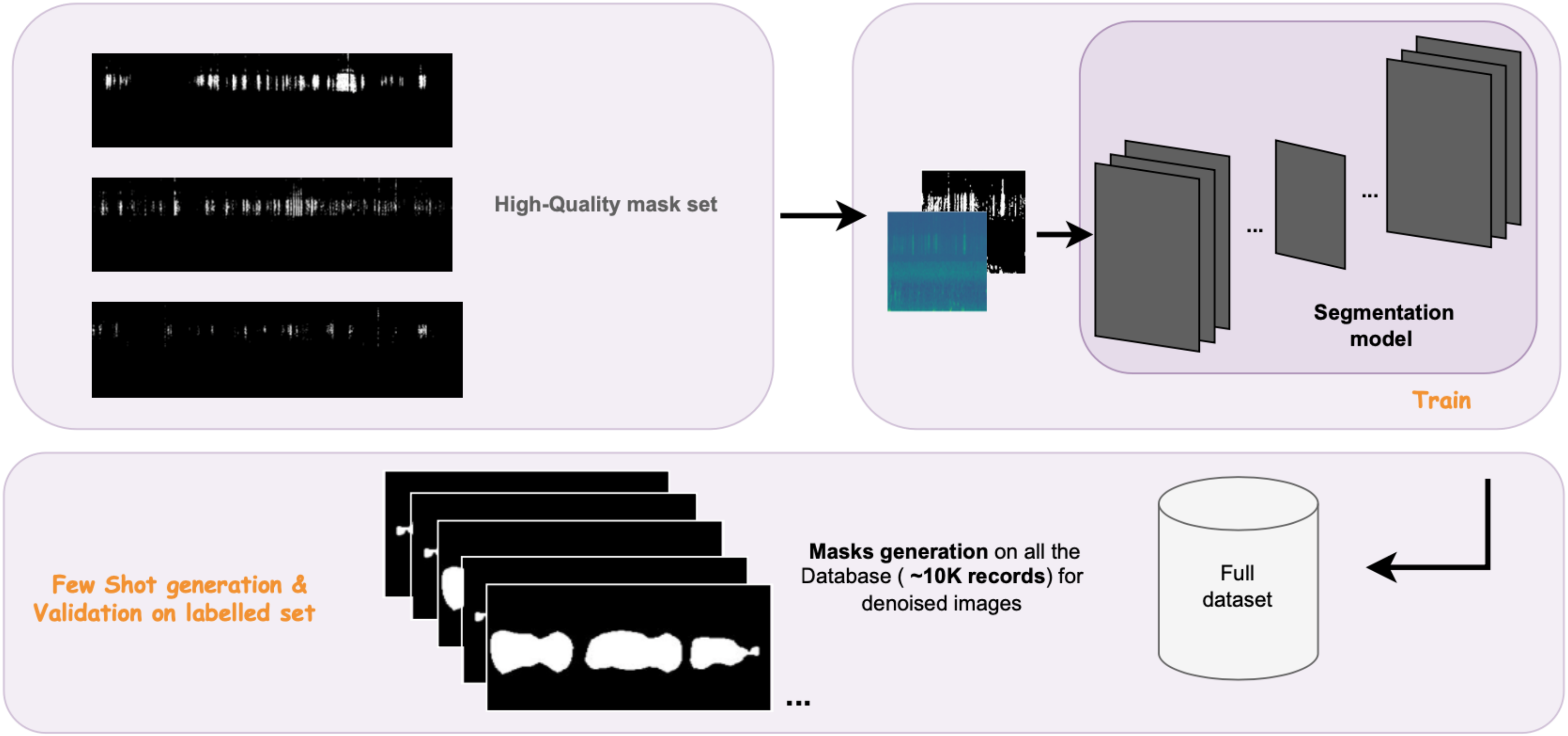
Time–frequency representation of raw acoustic signals



STEP 1 : From audio to spectrogram representation

Framework description

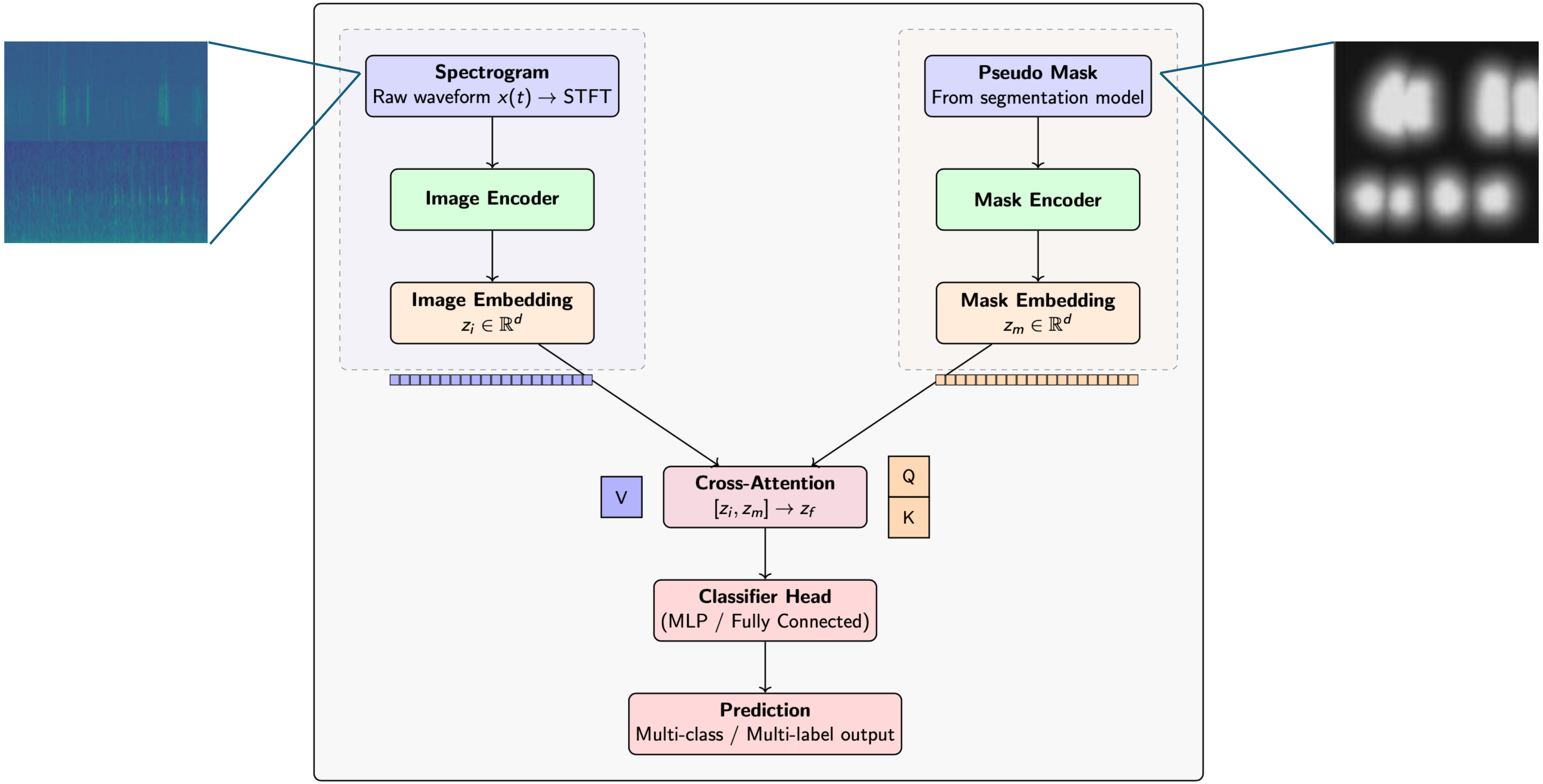
Few-shot learning strategy for segmenting vocalization regions of interest



STEP 2 : Generation of the masks

Framework description

Fusing spectrograms with generated masks to enrich signal representation



STEP 3 : Cross Attention Fusion and classification

Results

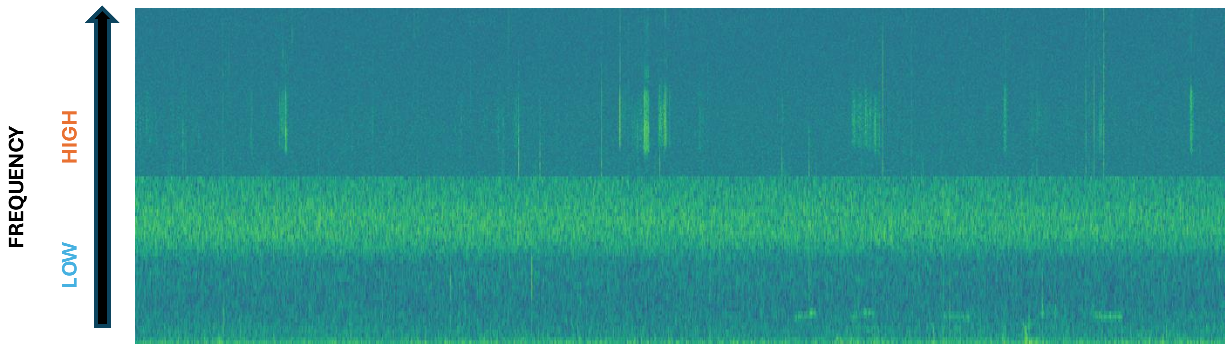
Comparison of baseline models with our approach

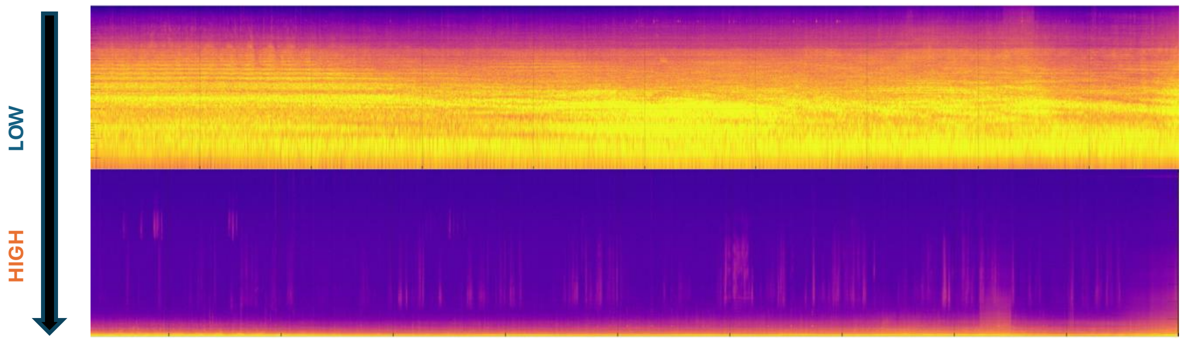
Model	Accuracy	F1 macro
Resnet50	0.588	0.562
ConvNext	0.625	0.591
ViT	0.788	0.787
Multi-step (Gen. Masks)	0.837	0.816
Multi-step (HQ Maks)	**0.897**	**0.890**

- Comparison of single-step baseline models and the proposed multi-step approach with cross-attention using either generated or a subset with high-quality masks.

Results

Generalization properties in OOD setting



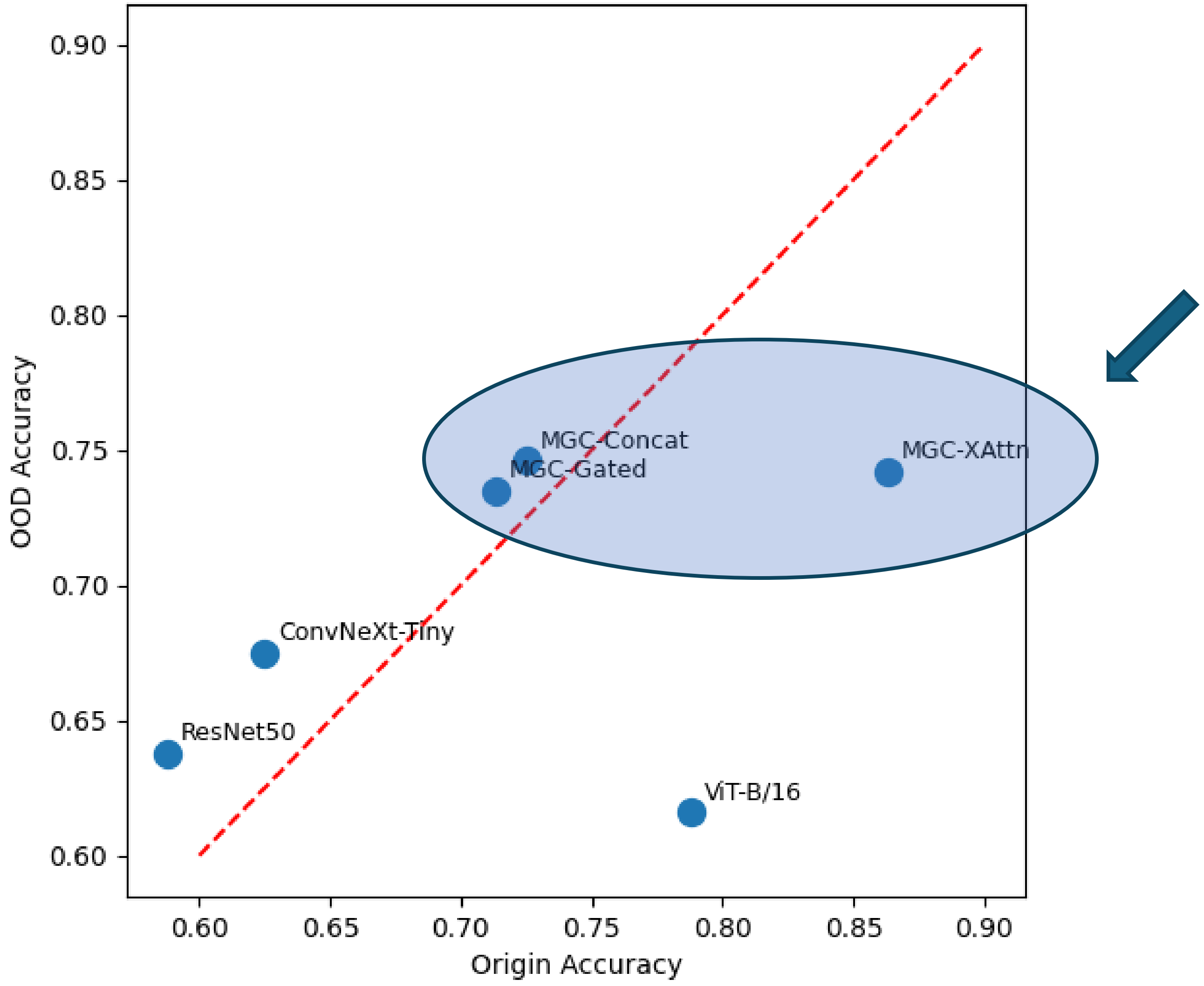


Aim of the Experiment :

- Assess robustness of models under distributional shift.
- Compare performance on original vs. OOD (out-of-distribution) data generated by alternative spectrogram transformations.

Results

Generalization properties in OOD setting



Mask guided classifiers achieve the best performance on both datasets

Conclusion

- Complexity of marine mammal vocalizations: Overlapping calls and variable noise make automated analysis challenging.
- Effectiveness of multi-step framework: Despite complexity, the proposed multi-step approach reduces ambient noise impact and improves classification robustness.
- Promising directions: Improving mask quality and integrating models that process raw audio can further enhance signal representation.

UQAR

Research Chair in Artificial Intelligence for Supply Chain Management (AI4SCM)












Lévis · XST Xpert Solutions Technologiques · mitacs · Québec · IVADO · ISMER UQAR