Review 1

PC member: Kilian Preuss
Time: Nov 20, 22:26 UTC

Review

This extended abstract presents a method that uses LoRA, a parameter-efficient fine-tuning technique, to adapt a pretrained ViT model for the classification and detection of whale calls. The performance of this method is then evaluated on multiple metrics and compared to other methods used in the BioDCASE Challenge.

The method seems to be well grounded in the current deep learning paradigm. In fact, LoRA is a widely used fine-tuning approach, and ViTs tend to outperform CNNs in high-data regimes. The wide variety of metrics used provides substantial insight into the strengths and weaknesses of the method. Furthermore, comparing the method with other approaches highlights its advantages and contributions relative to existing techniques.

On the other hand, it is not specified what kind of images the ViT was pretrained on, nor how the knowledge learned from these images could help classify spectrograms of whale calls. Moreover, it remains unclear why converting the audio into a spectrogram and training an image-based model is a better approach than fine-tuning an audio foundation model directly on the raw audio.

The extended abstract would benefit from an introduction that provides more context about related work, situates the study within the existing literature, motivates the chosen methods, and explains how the proposed approach differs from previous methods and what makes it novel. Apart from that, major concepts such as ViT and LoRA are well explained, and the document provides many details regarding the dataset and the evaluation metrics. The organization of the paragraphs follows a logical flow, making the document easy to read and understand.

Reviewer's confidence

2: Partly, I may be missing some concepts or elements of the state-of-the art, but I got the main idea

Usage of LLM

1: No, not at all

Confidential remarks for the program committee

(None provided)