Testing the Engine Model on Real-World Data

Jessica Stinson
224576666

Introduction

Founded in 2022, Project Echo is dedicated to developing a sound classification model to facilitate the monitoring of endangered species in the Otways, Victoria. The current model can classify 118 species within the area, with ongoing efforts to expand the database to include all regional species. In the long term, the project aims to broaden the model's capabilities to recognise a diverse range of animal vocalisations and environmental sounds. Project Echo provides a non-intrusive approach to tracking both endangered species and their predators to support the work of animal conservationists. When challenged with test datasets, the model performs extremely well, with the most recent testing consistently yielding an accuracy greater than 95%. While this provides a promising outlook for the future success of Project Echo, the model is yet to be tested on real-world data collected using microphones in the field. This report describes and critiques initial real-world testing of the model with recommendations for future work.

Testing the Engine Model

Prior to exposing the model to real-world data, it was tested on the test dataset to ensure the reliability of results when working on known data. Initial testing yielded an accuracy of 98.68%, so real-world testing proceeded. The model was tested on an hour-long audio recording from the Otways National Park. The recording was split into clips of 2-second intervals and converted into spectrograms for classification. While the real-world data available for testing is extensive, memory limitations impacted the volume of data that the testing pipeline could process.

The audio currently available for real-world testing is not labelled, so the true class of vocalisation is not currently known, and metrics such as accuracy and precision cannot be assessed. Instead, a subset of the audio clips was manually analysed by comparing the audio clips to the testing clips for identified species.

Discussion

Analysis of the real-world test results revealed several noteworthy patterns and areas for improvement in future testing. In a subset of the audio clips, the vocalisations matched those of the species identified by the model, indicating some successful classifications, however, most samples appeared to be misclassified. In many cases, the vocalisations in the real-world recordings were similar to the test dataset, but most likely originated from different species. Given that the model was trained on only 118 species from the Otways region, this was not entirely unexpected. When encountering unfamiliar species, the model is likely to assign the closest known match from its training dataset.

Another significant issue encountered was the degradation in audio quality caused by downloading the dataset via Microsoft Teams. The reduced quality made it difficult to discern vocalisations clearly, likely contributing to false positives. For instance, the model occasionally identified the species *Dama dama* (European fallow deer) in clips that, upon listening, seemed to contain no distinct animal vocalisations. These misclassifications suggest that the model may have been interpreting background noise as species-specific calls. Currently, the model lacks the functionality to filter or skip audio clips without vocalisations, which increases the risk of incorrect classifications. This is particularly problematic in conservation contexts, where misidentification could hinder the accurate monitoring of endangered species.

While the training and testing datasets consisted of curated, single-species audio clips, real-world data typically includes multiple overlapping vocalisations from various species within the same recording. The current model is designed to select one species per clip, which led to a bias in classifications during real-world testing. Certain calls were identified more frequently than others, potentially overshadowing less common or quieter species. This is a critical limitation, especially regarding underrepresented or endangered species, which may be consistently overlooked by the model.

Conclusion

To enhance the model's utility in conservation and ecological monitoring, several improvements should be prioritised. The training dataset must continue to grow to better represent the full population of the Otways region. With growing interest in the testing of the engine model on real-world data, it is recommended that alternative methods be considered for sharing data samples for real-world testing, such as uploading the data to Google Cloud for sharing. Future efforts may also consider integrating functionality to skip non-informative audio segments and classify multiple species in a single audio file.