ABSTRACT FOR DCASE 2023 CHALLENGE SUBMISSION

FEW-SHOT BIOACOUSTIC EVENT DETECTION USING BEATS

Technical Report

Femke Gelderblom[1], *Benjamin Cretois*[2], *Pål Johnsen*[3], *Filippo Remonato*[3],

[1] Acoustics, SINTEF Digital, Trondheim, Norway
[2] Environmental Data, Norwegian Institute for Nature Research, Trondheim, Norway
[3] Mathematics and Cybernetics, SINTEF Digital, Trondheim, Norway

Our method for the DCASE Challenge 2023 combines BEATs with Prototypical Networks. BEATs, standing for Bidirectional Encoder representation from Audio Transformers, is a newly-released architecture by Microsoft for audio tokenisation and classification. BEATs combines a tokenizer and a semi-supervised audio classifier which learn from each other to improve the classification of audio samples. Prototypical Networks, instead, can be briefly described as a neural network-based clustering algorithm. Somewhat resembling a K-means clustering, Prototypical Networks classify samples based on their distance from the classes' prototypes (what would be the centroids in a K-means setting). Since the prototypes are constructed from a small set of examples from each class, called the support set, Prototypical Networks are well suited to handle few-shot learning settings like the DCASE Challenge. In our method, we combine the two by using BEATs as a feature extractor, constructing informative features which are used by the Prototypical Network to perform the prototypes' construction and subsequent classification of test audio samples. We obtain a F1 score of 0.36 on the validation dataset.