We present a learning-based method for detecting whistles of toothed whales from underwater hydrophone recordings. Our method represents audio signals as time-frequency spectrograms and employs the Fully Convolution Network (FCN) to estimate for each spectrogram a map of contour confidences that are used for extracting discrete whistle contours. To avoid the expensive efforts of annotating whistle contours, we develop a data synthesis approach to generate spectrogram-contour pairs using spectrogram of background environment and a small set of whistle contours. Our study suggests that the deep contour model can be effectively learned from these synthesized samples. However, it is costly and unnecessary to synthesize equal amounts of samples for each spectrogram or contour. Instead, we present an alternative learning algorithm that synthesizes samples only for those spectrogram or contours that are not well modeled by the current network, measured by recall rates of contour points for each spectrogram contour sample. This recall-guided learning algorithm can adaptively synthesize difficult samples to boost learning effectiveness. We applied the proposed method to the public DCLDE2011 dataset to extract whistle contours. Results show that our method can improve state-of-the-art method up to 21.9 percent in terms of F-score for multiple odontocete species.

Pu Li, Marie Roch, and Xiaobao Liu

This research is supported by the Computational Science Research Center (CSRC) at San Diego State University.