深層学習による長時間録音からの高精度な鳥類音声の自動抽出

○水村春香・安田泰輔・松山美恵・塚田安弘・瀧口千恵子（山梨富士山研）

1. はじめに

音響による鳥類相モニタリングは，その効率の良さや網羅性から世界的に活用が進んでいる。また，生物多様性指標の効率的な観測手法の一つとして，社会的需要も増している。しかし，膨大な録音の中から，網羅的に鳥類の音声を抽出することは難しく，音響モニタリングの拡充におけるハードルとなっている。本報告では鳥類の音声抽出においてこれまで使用例のない Semantic Segmentation（SS）を用いて，従来よりも少ない教師データから富士山麓の鳥類の音声のみを効率的かつ正確に抽出する手法の開発を試みた結果を報告する。

2. 方法

調査は，山梨県富士山科学研究所の森に IC レコーダーを設置し，2023 年 6 月 9 日から 28 日のうち 6 日間，各日 1 時間の録音を無作為に抽出し計 6 時間の録音を解析対象とした。音声データをスペクトログラム画像へ変換した画像データを使用し，これに対して”鳥類の音声（複数種含む）”と”鳥類の音声ではない（無音も含む）”の 2 クラスで色分けし，教師データを作成した。SS モデルとして DeepLabv3+を採用し，エンコーダーとして ResNet18 を用いた。4 時間の録音からトレーニング，バリデーション，テストデータそれぞれ 1248 枚，416 枚，416 枚と 6:2:2 となるよう分割した。2 時間の録音は性能評価（適合率，再現率，F1 Score，マクロ平均の算出）に用いた。

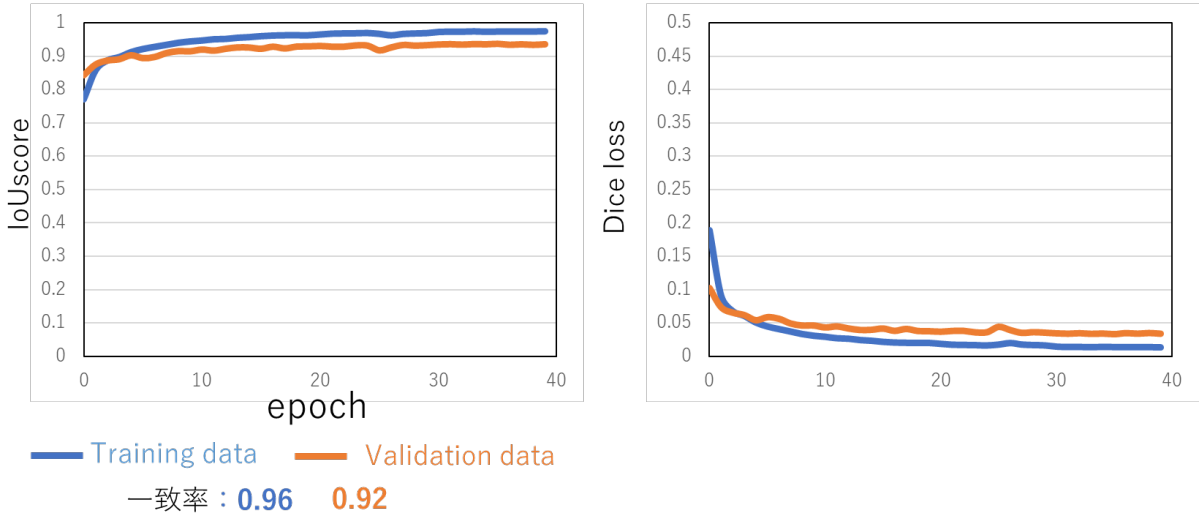
3. 結果と考察

トレーニング及びバリデーションに基づき構築したモデルのテストの結果（図），複数種の鳥類の音声が含まれていても IoU スコア（一致率）は 0.936，損失関数は 0.04 と高い精度で鳥類の音声を抽出可能であり，過学習も起きていないことが示された。F1 Score は 0.84，0.81 となったが，音源によっては再現率が 0.6 と低くなっており，偽陰性がやや高いことがわかった。音圧が弱い部分で誤判定が生じている可能性があり，音圧への閾値を決めるなどの対応で精度の向上が期待される。今後は異なる環境での適用や，特定の種の判別まで含めて自動抽出を進める予定である。

引用文献

Chen *et al*. (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv e-prints. arXiv preprint arXiv:1802.02611, 10.

Stowell, D. (2022) Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*. Vol. 21, 10:e13152.



図．トレーニング，バリデーションの結果．

左図：トレーニング及びバリデーションにより，epoch=10 ほどで IoU スコア 0.9 以上となった．右図：Dice loss も epoch=10 ほどで 0.05 を下回った．