

深層学習モデルを用いた 音声による鳥の鳴き声の検索

バードリサーチ鳥類学大会2023

加藤 義清(バードリサーチ会員)
X (旧Twitter): @ykt_bird

はじめに

社会的意義

自然保護活動の第一歩として生態系の状況を知ることは欠かせない

安価で入手が容易なセンサーの普及により利用可能な生態学データは大幅に増えており、生態系のさらなる理解促進が期待される

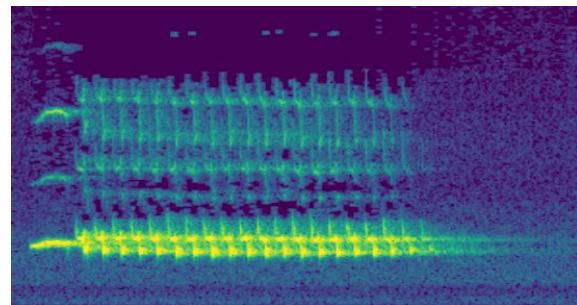
生態学で用いられてきた従来のデータ解析手法は大量のデータ処理に対応できていない。近年発展しているデータ解析、機械学習の技術の応用が課題解決の一助となると目されている^[1]

本発表では蓄積された鳥類の鳴き声の音声データの活用の方法として音声による鳴き声の検索技術を検討する

個人的動機

鳥の鳴き声から種を同定したい

種が知りたいだけでなく、同じ種の鳴き声を聴き比べたい



コマドリのさえぎり

※「CD鳴き声ガイド日本の野鳥」収録の音声より作成

[1] Tuia et al.: Perspectives in machine learning for wildlife conservation. Nature Communications (2022)13:792. <http://doi.org/10.1038/s41467-022-27980-y>

音声による鳥の鳴き声の検索

検索とは

データベースに登録されたデータのうち、問合せ(クエリ)で指定された検索条件に適合したものを検索結果として応答するプロセス

例

1) 図書館の蔵書検索

検索対象 書籍データ(図書館の蔵書)

問合せ 本のタイトル、著者名等

結果 問合せ内容に合致した書籍の一覧

2) 検索エンジン

検索対象 ウェブページ

問合せ キーワード

結果 キーワードに関連するページの一覧

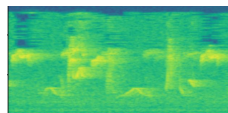
音声による鳥の鳴き声の検索

検索対象 鳥の鳴き声の音声データ

問合せ 音声データ

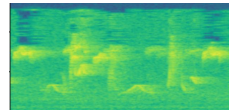
結果 問合せに類似した音声データと付随する属性
(種名、録音日時、場所、録音者)の一覧

問合せ

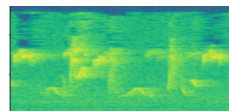


種名 ?

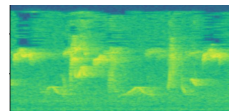
検索結果



シジュウカラ
2013/05/03, X市



エナガ
2017/06/10, Y村



ヤマガラ
2015/04/21, Z町

埋込ベクトル空間における近傍探索による検索

- ・ 深層学習モデルを利用して音声データを埋め込みベクトル空間に写像
- ・ 埋込ベクトル空間の性質を利用して類似音声を検索

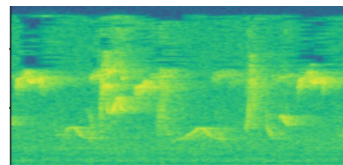
埋込ベクトル



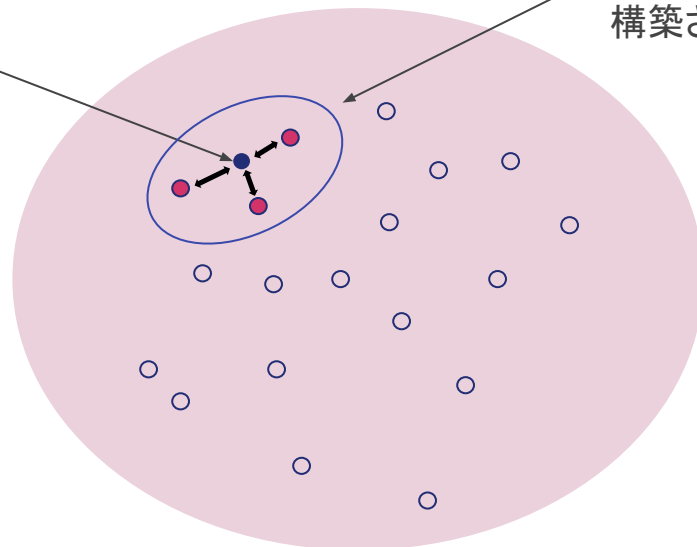
深層学習モデル



音声データ



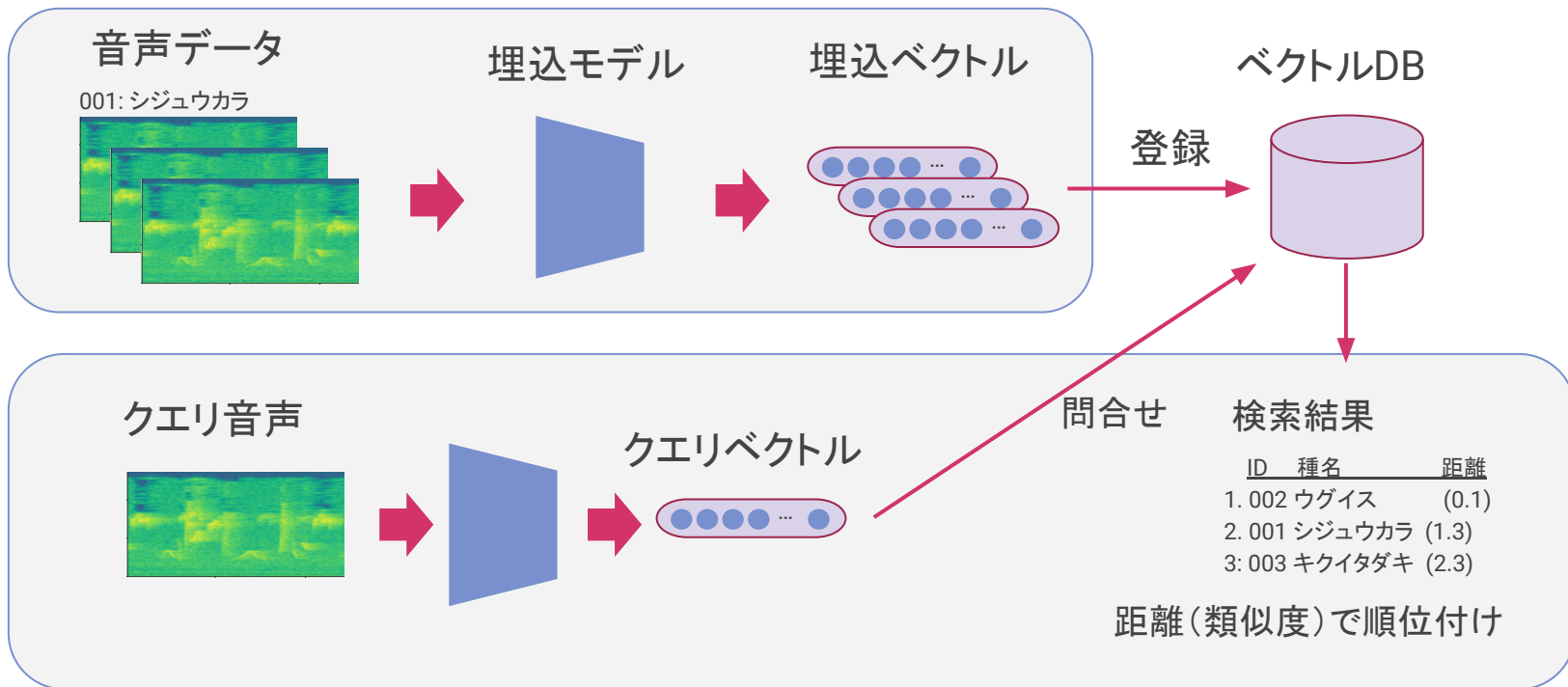
ベクトル空間は類似する音声
同士の距離が小さくなるように
構築されている



埋込ベクトル空間

システム構成

インデックス作成



深層学習モデル

鳥の鳴き声データで学習された深層学習モデルが様々な生物音響学タスクに応用できることが知られている^[2]

深層学習モデルにはGoogle Bird Vocalization Classifier^[3] (以下、GBVC) を使用

アーキテクチャ: EfficientNet-B1^[4]

学習データ

Xeno-Canto^[5a]の鳥の鳴き声の音声データ

但し、以下のものは除外

- ・ライセンスで改変禁止指定されたデータ
- ・IUCNレッドリストに登録されている種のデータ

学習方法: 音声データから種ラベルを予測

入力: サンプリング周波数32kHzで5秒分の音声データ

出力: 種予測値 (eBird種ID 10932種)
埋込ベクトル (1280次元)

[2] Ghani et al. (2023) <https://arxiv.org/abs/2307.06292>

[3] <https://www.kaggle.com/models/google/bird-vocalization-classifier>

[4] Tan and Le (2019) <https://arxiv.org/abs/1905.11946>

[5] <https://xeno-canto.org/>

Xeno-cantoについて

サイト: <https://xeno-canto.org/>

野生生物音声録音の普及、野生生物音声データへのアクセス提供と理解促進を目的とした野生生物の音声共有サイト

現在は主に鳥類、バッタ類、コウモリ類を対象としている

鳥類の音声データの概要 (2023年12月現在)

種数: 約10,000種

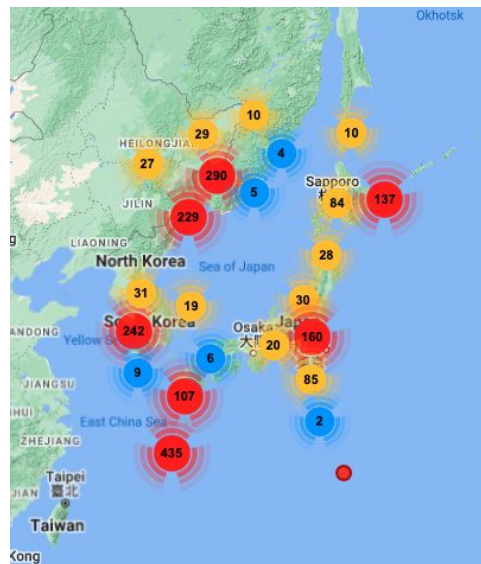
音声データ: 約80万件

総録音時間: 11600時間

APIでプログラムからデータにアクセス可能

<https://xeno-canto.org/explore/api>

日本周辺の音声データも収録されている



ベクトルデータベース

埋込ベクトルを付随するデータ項目とともに管理し、高速で類似検索を可能にする

本発表ではベクトルDBとしてmilvusを使用

サイト: <https://milvus.io/>

類似度としてユークリッド距離、内積、コサイン類似度を利用可能

複数の索引タイプが利用可能で、類似検索の速度と正確性のバランスを調整可能

性能評価実験

方法

- 1) 種ラベルが付与された音声データをベクトルDBに登録(検索対象データ)
 - ・5秒間のフレームを1秒間隔で切り出して登録
 - ・類似度尺度としてユークリッド距離を使用
 - ・バードリサーチ 鳴き声図鑑^[6]の音声データ (BR)
400種1228件
- 2) 検索対象データとは異なる種ラベルが付与された音声データからクエリ音声集合を用意
 - ・CD鳴き声ガイド 日本の野鳥 (BSOJ)
377件 (1種につき1件5秒分の音声を切り出し)

3) 各クエリ音声につき、検索結果の上位10件について以下の指標で評価する

- 上位k件適合率 (Precision@k)の平均
- 平均適合率の平均 (Mean Average Precision)

[6] https://www.bird-research.jp/1_shiryo/nakigoe.html

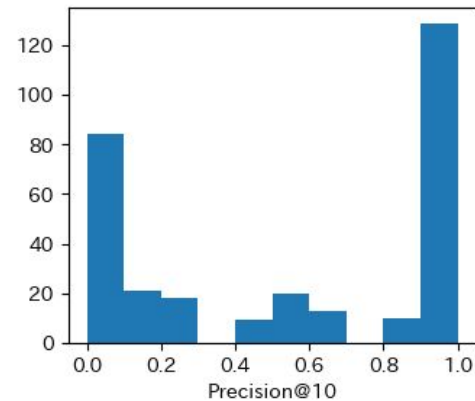
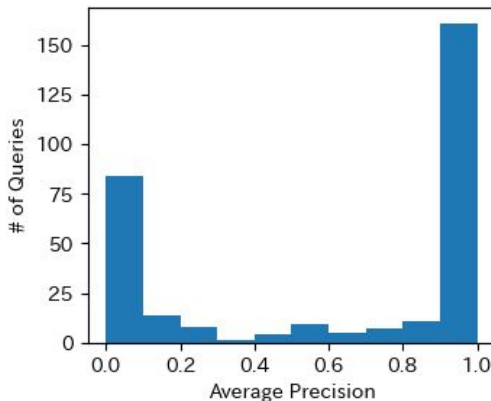
結果

結果は下表のようになった。

クエリ集合に用いたBSOJデータには、BRデータに含まれない73種の音声データが含まれていたため、それらのデータを除いて評価した結果も併記した。

BSOJにのみ含まれる種の例: ミヤマカケス、サカツラガン、スズガモ、トモエガモ、リュウキュウウグイス、アホウドリ等

クエリ毎(つまり種毎)の性能の分布を右図に示す。
検索がうまくいっている場合とうまくいっていない場合に二極化している様子が見て取れる。



評価条件	MAP	P@1	P@5	P@10
クエリ集合全体で評価した場合	0.500	0.496	0.465	0.440
クエリ集合から検索対象に存在しない種を除いて評価した場合	0.620	0.615	0.577	0.546

成功例

117種で検索結果上位10件が全て適合となった。

カイツブリ オオハクチョウ カリガネ ヒシクイ ヨシガモ オナガガモ ヒドリガモ コオリガモ ケイマフリ コアジサシ サンカノゴイ ミゾゴイ コウノトリ ヒクイナ シマクイナ ヤンバルクイナ ケリ アオアシシギ キアシシギ ダイシャクシギ	チュウシャクシギ オオジシギ セイタカシギ ソリハシセイタカシギ カンムリワシ トビ オオタカ ツミ ワシミミズク シマフクロウ フクロウ コノハズク リュウキュウコノハズク オオコノハズク キンメフクロウ アオバズク ライチョウ エゾライチョウ ミフウズラ ウズラ	コジュケイ カッコウ ツツドリ セグロカッコウ ホトトギス ジュウイチ ヨタカ ヒメアマツバメ アマツバメ アカショウビン カワセミ アリスイ オオアカゲラ ヤイロチョウ ヒバリ イワツバメ ツバメ ハクセキレイ ビンズイ タヒバリ	リュウキュウサンショウクイ ヒレンジャク キレンジャク カワガラス ミソサザイ イワヒバリ コマドリ シマゴマ ルリビタキ マミジロ トラツグミ アカハラ アカッコ シロハラ ヤブサメ オオセッカ キクイタダキ ウチヤマセンニュウ マキノセンニュウ エゾセンニュウ	ヤナギムシクイ カラフトムジセッカ オオムシクイ メボソムシクイ エゾムシクイ センダイムシクイ イジマムシクイ マミジロキビタキ オジロビタキ ニシオジロビタキ サンコウチョウ コガラ ヒガラ シジュウカラ エナガ キバシリ メジロ カシラダカ アオジ シマアオジ	クロジ カワラヒワ ベニヒワ ギンザンマシコ イスカ ベニマシコ ニュウナイスズメ ハッカチョウ カケス ルリカケス オナガ ホシガラス コクマルガラス ハシブトガラス ワタリガラス ホンセイインコ ソウシチョウ
--	--	---	---	--	--

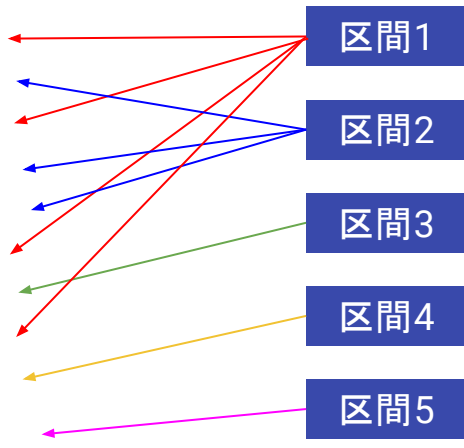
失敗例

84種で検索結果上位10件が全て不適合となった。

カンムリカイツブリ アカエリカイツブリ ハジロカイツブリ ウミウ コブハクチョウ コハクチョウ コクガン ハクガン ハシビロガモ コガモ オシドリ キンクロハジロ カワアイサ カンムリウミスズメ オオミズナギドリ コシジロウミツバメ セグロカモメ ウミネコ カモメ ズグロカモメ アジサシ ヨシゴイ	ゴイサギ コサギ ダイサギ タンチョウ ソデグロヅル バン シロハラクイナ イカルチドリ メダイチドリ ムナグロ トウネン ハマシギ ツバメチドリ イソシギ ソリハシシギ キョウジョシギ オオソリハシシギ ホウロクシギ ミサゴ オジロワシ オオワシ イヌワシ	ハチクマ サシバ チュウヒ ハヤブサ ヤマドリ キジバト ヤマセミ ブッポウソウ ヤツガシラ クマゲラ アカゲラ ノグチゲラ コアカゲラ リュウキュウツバメ セグロセキレイ シロガシラ ヒヨドリ チゴモズ アカモズ ヤマヒバリ イソヒヨドリ	クロウタドリ セッカ コヨシキリ チフチャフ オオルリ サメビタキ エゾビタキ メグロ ホオアカ ミヤマホオジロ シベリアアオジ オオマシコ ウソ ムクドリ コムクドリ ギンムクドリ ミヤマガラス インドクジャク シマキンパラ
--	--	--	---

成功事例 キクイタダキ

- 1 キクイタダキ (437:21)
- 2 キクイタダキ (437:8)
- 3 キクイタダキ (437:20)
- 4 キクイタダキ (437:7)
- 5 キクイタダキ (437:6)
- 6 キクイタダキ (437:19)
- 7 キクイタダキ (430:28)
- 8 キクイタダキ (437:22)
- 9 キクイタダキ (430:14)
- 10 キクイタダキ (437:15)

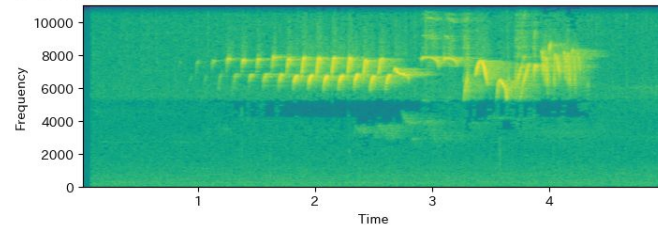


フレーム内の音声位置がずれていても適合することを確認できた

2つの音声データ(437と430)の重複する区間に適合
→適合率が過大評価されている可能性

クエリ音声

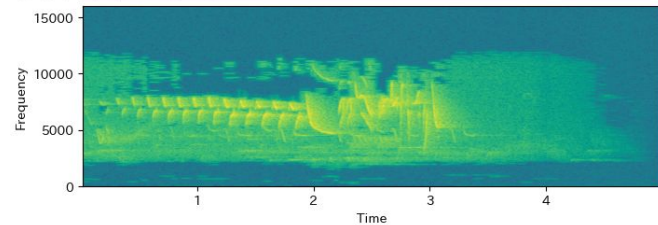
キクイタダキ (Regulus regulus)



検索結果

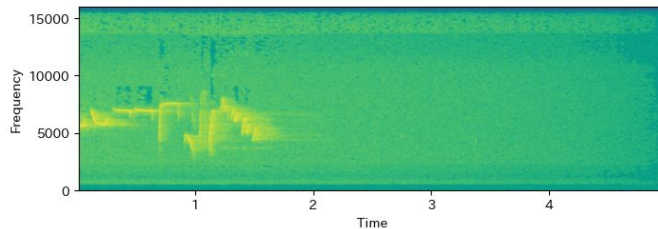
437:21

キクイタダキ (Regulus regulus)



430:28

キクイタダキ (Regulus regulus)



失敗事例1 カムリカイツブリ

検索結果

- 1 インドクジャク (74:35)
- 2 インドクジャク (74:2)
- 3 インドクジャク (74:34)
- 4 インドクジャク (74:20)
- 5 ハシブトガラス (667:10)

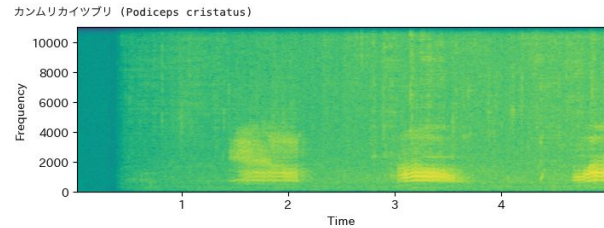
インドクジャクもハシブトガラスもカムリカイツブリと似た声質

→類似音声として検索されたか

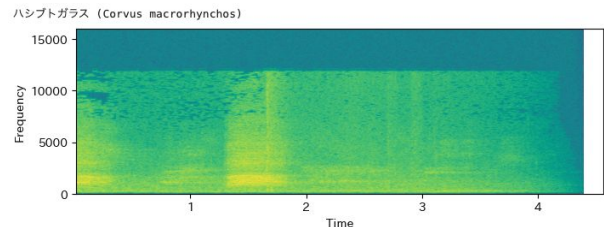
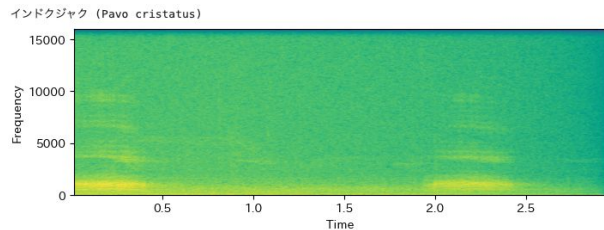
BRデータには3件のカムリカイツブリのデータが含まれるがいずれもクエリ音声には似ていない

→種内の声質の分散が影響か

クエリ音声

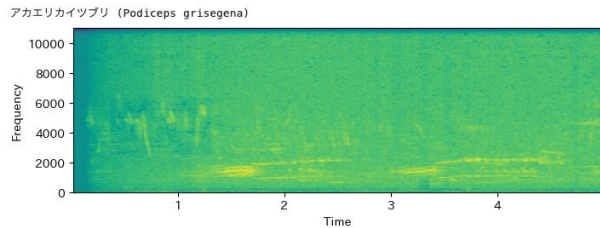


検索結果



失敗事例2 アカエリカイツブリ

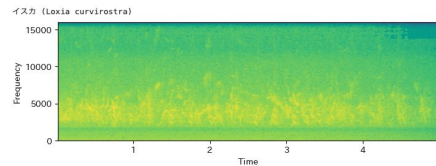
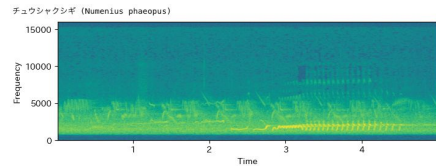
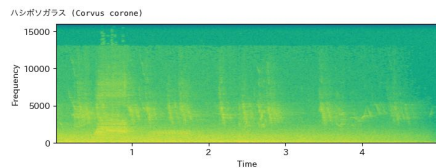
クエリ音声



検索結果

- 1 ハシボソガラス (679:11)
- 2 チュウシャクシギ (1041:6)
- 3 イスカ (136:16)
- 4 ハツカチョウ (43:4)
- 5 チュウシャクシギ (1041:7)

検索結果



クエリ音声および上位の検索結果音声はラベル種以外の鳥の鳴き声が顕著に混ざっている

失敗事例3 ウミウ

検索結果

- 1 ヒメウ (953:11)
- 2 ミユビシギ (1028:6)
- 3 ヒメウ (953:16)
- 4 ヒメウ (953:12)
- 5 ヒメウ (953:10)
- 6 カワウ (949:20)
- 7 ヒメウ (953:17)
- 8 カワウ (948:18)

近縁種との取り違えか

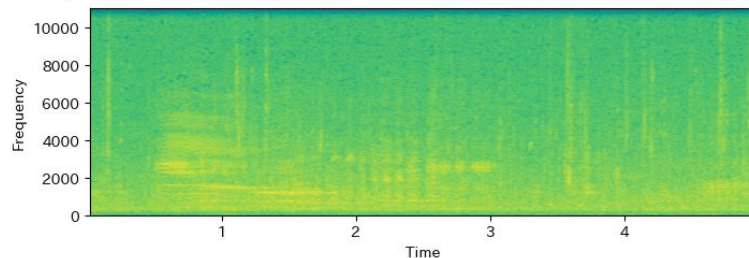
ミユビシギの音声は波の音しか録音されていないかった
→環境音の影響か

BRデータにはウミウは1件のみで、音声はクエリ音声に類似しない

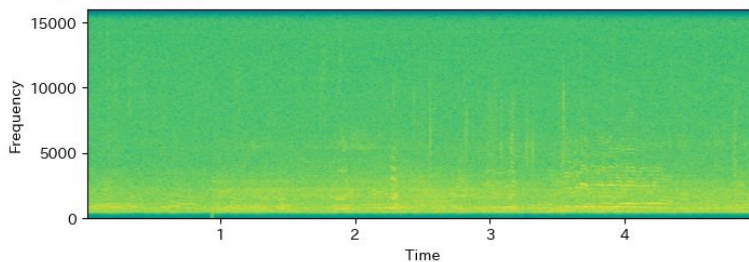
クエリ音声

検索結果

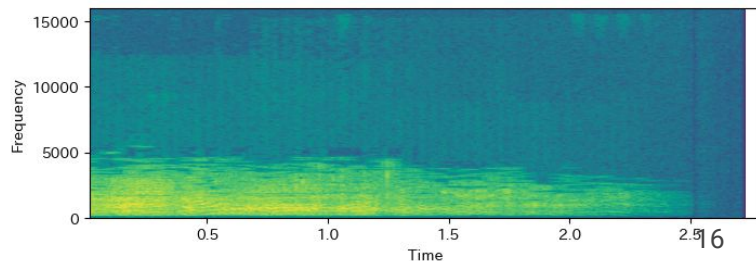
ウミウ (*Phalacrocorax capillatus*)



ヒメウ (*Urile pelagicus*)



カワウ (*Phalacrocorax carbo*)



考察

性能評価について

直接比較の対象がないので解釈は難しいが、今回用いた深層学習モデルのGBVCは、その性能評価において8個のデータセットに対し以下の数値を示している。

MAP [0.33, 0.67] (Mean=0.58; Var=0.011)

Top-1 [0.48, 0.86] (Mean=0.62; Var=0.013)

これと比較して、本実験は種同定に関して遜色のない性能を示していると言える。

評価方法について

重複する複数の区間も適合として扱っているので適合率が過大評価されている可能性がある。重複区間の扱いについて要検討
(例えば連続する区間は一つの検索結果として統合するなど)

失敗事例分析

少数の失敗事例の分析から、次のような要因で失敗する可能性があると考えられる

- 1) 種間の鳴き声の声質が似ている
- 2) 種内の声質の分散が大きい
- 3) 複数種の鳴き声が混じっている
- 4) 環境音が影響を与える

検索は種同定のみが目的ではなく、鳴き声が類似する種の実分析や、種内の鳴き声の分散の実分析に利用可能性がある点において、1と2に関しては必ずしも問題とはいえない。

3や4については、今後対応の検討が必要であると考えられる。

今後の課題と展望

結果のさらなる分析

少数事例の分析にとどまっているので、より多くの事例を分析して知見を増やしたい

音声分離手法の適用

複数種の音声が混じっている場合に問題となることが確認された。鳥の鳴き声の音声分離について近年提案された手法^[7]があるので、その適用を検討したい

類似音声検索システムの構築

現在のプロトタイプをもとに、実利用可能なシステムの構築・運用を検討したい

[7] Denton, Wisdom & Hershey: Improving Bird Classification with Unsupervised Sound Separation, ICASSP 2022. <https://arxiv.org/abs/2110.03209>

まとめ

深層学習モデルを応用した音声による鳥の鳴き声の検索技術を紹介した

性能評価実験の結果、種同定性能は悪くはないが改善が必要なことも明らかになった

失敗事例の検討により、特に複数種の鳴き声が混ざる場合と、環境音が問題となることが分かった

鳥類の調査研究や、自然環境保護の啓発活動等への関連技術の応用に興味がある方は、ご連絡ください。

X (旧Twitter)アカウント: @ykt_bird