

画像識別による IC レコーダーを用いたヤンバルクイナの鳴き声自動検出 Automatic Detection of Gallirallus-Okinawae Using an IC Recorder by Image Identification

森下 功啓[†]
Katsuhiko Morishita

1. はじめに

環境調査の一環として世界中で野鳥の生息調査が行われており、現時点では定点での目視や聞き耳を立てることでの観察、鳴き声をスピーカーで流してその反応を確認する調査が行われている。これらは調査対象が縄張りを持つ種や警戒心の弱い種に限られやすいという課題と、調査員に高い技能と労力を要するという課題がある。

そこで、環境中で自動録音した音源から野鳥の種を自動識別する手法の研究が行われている。これが実現すれば、環境中への IC レコーダーの設置とその音源のデータ処理によって広い地域での同時調査が可能となり、より高密度な環境調査が行える。ただし、先行研究では野鳥とわかっている短い音源の識別や、特定の種に限った識別が行われており、幅広い種の調査はまだ実現できていない^[1]。

本研究では、環境音に含まれる多種の野鳥を識別することを目的とし、複数の野鳥だけではなく昆虫やカエル等や人工音を含めて VGG16 を転移学習させた。本論文では、特に保護対象として優先度の高いヤンバルクイナについて取り組んだ内容を報告する。

2. 野鳥の特徴

2.1 野鳥の鳴き声の特徴

野鳥は鳴管と呼ばれる発声器官を持ち、2 つの周波数を独立に出すことで複雑な発声が可能である。鳴き方は、ピー・ピー・ピーの様に比較的単純な音や、それを繰り返すか否か、チ・チと短い音を出す、笛を吹くような音色を繰り返す、などがあり、複雑なフレーズを絶え間なく鳴き続けるヒバリの様な種もある。声の周波数は 100 Hz 前後～10 kHz ほどである。

鳴き方には繁殖期によく聞かれる囀りと、それ以外の地鳴きがある。囀りは種によって大きく異なることが多いが、地鳴きは地味でプロでも区別が難しいことが多い。加えて、それぞれの種が縄張りの主張や警戒などの数種類の鳴き方を持っている。また、囀りや縄張りの主張は季節性であるので、鳴き始めとそれ以降では習熟度に差があり、鳴き方が徐々に変わる。また、鳴き方には地方性がある。

2.2 ヤンバルクイナの特徴

ヤンバルクイナは沖縄北部の山原地域に生息する日本固有種である。外来種のマングースにより生息数を減らしてきたが、最近のマングース駆除活動により 1600 羽程度まで回復したとみられている。

鳴き声は主にケケケケケという連続的な声で、周波数

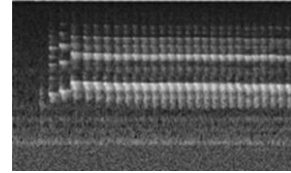


図 1 作成したスペクトログラム画像の例

は 1.5~4 kHz 付近である。この鳴き声は数秒～30 秒ほど継続し、徐々に小さくなる。その他、警戒音などを含めると 5~6 種類ほどの鳴き方がある^[2]。

なお、ヤンバルクイナは頻繁に鳴く野鳥ではなく、鳴き声は録音時間の 0.4 % 程を占めるに過ぎない。山原地域で録音した音源の大部分は無音やほかの野鳥や虫の鳴き声や海岸から伝わる波浪の音が占めている。

3. 提案手法

複数の野鳥を識別するために、音源から 5 秒ごとにスペクトログラムを作成し、これを画像識別でよく使われる VGG16 で転移学習を行った。なお、複数の音が重なることを考慮し、学習モデルには Softmax 層を付けていない。

3.1 識別対象種

主な識別対象種はヤンバルクイナに加え、同じ環境中で鳴く野鳥や虫を中心に、ウグイス・カラス・リュウキュウコノハズク、リュウキュウアカショウビン、ホントウアカヒゲ、ヒヨドリ、モズ、ヒガラ、スズメ、ヤイロチョウ、その他の市販 CD に収録された野鳥、リュウキュウサワマツムシ、リュウキュウカネタタキ、マダラコオロギ、ケラ、オオシマゼミ、ニイニイゼミ、クマゼミ、山原地域で鳴く虫 2 種、その他の市販 CD に収録された鳴く虫、を識別対象とした。その他、市販 CD 収録のカエル、ネコ、車の各種音、救急車、踏切、雨、風、小川、静寂、エアコンの室外機、金属製の扉、テーピングの音を識別対象とした。

3.2 環境音の収集

主に Panasonic 社製 RR-XS460/470 を用いて、沖縄県山原地域でタイマー録音を行った。録音は 2017 年から行い、2018 年以降は毎年延べ 500 時間超の音源を得ている。なお、Panasonic 社製 RR-XP007 や TASCAM 社製 DR-05 を用いて短時間の録音も行った。その他、熊本県八代市を中心に森林内での録音を行うとともに、東京大学演習林で収録されたサイバーフォレストの音源と、バードリサーチ様を通して頂いた音源を収集した。また、一部は YouTube の音源を利用した。

3.3 教師データの作成

収集した音源から作成した画像の例を図 1 に示す。縦軸は周波数を表し、下端 0 Hz 上端 22 kHz のメルスケールで

[†]熊本高等専門学校建築社会デザイン工学科
Department of Architecture and Civil Engineering, National
Institute of Technology, Kumamoto College

ある。横軸は時間を表し、サイズ 120×200 pixel のモノクロ画像である。なお、小さい音も画像化するために、スペクトログラムを作成後に対数を掛け、範囲を 0~255 に正規化した。さらに、風の音などの雑音がある状態と静寂の差を付けるために、最上部の画素に音声の振幅情報を埋め込んだ。また、必要に応じて周波数フィルターを掛け、不要な雑音部分には同じ音圧のノイズを挿入した。

作成した画像は種類ごとにクラスに分けて教師データとした。また、複数のクラスにまたがる画像も加えた。なお、教師データの取捨選択においては、なるだけ多様な鳴き声を含めた。ただし、ウグイスの囀りと谷渡りなどの鳴き方が明らかに異なる上に画像数が確保可能なものについては別のクラスとした。

教師データの画像数は全部で 67,661 枚となった。特にヤンバルクイナは多く作成し、単独のラベルが付いた画像が 1,268 枚となった。

3.4 学習

教師データを Python 3.8 と TensorFlow 2.5.0 をバックエンドとした Keras 2.4.3 フレームワークを用いて学習させた。900 epoch までは追加したユニットに対する学習を行い、901~1600 epoch は第 6 層以降の学習を行った。学習係数はそれぞれ 0.001 と 0.0001、バッチサイズは 220 とした。

Optimizer は SGD、損失関数は Binary Crossentropy を用いた。

教師データは 95:5 で学習用と検証用に分割し、学習を行った。学習中は、教師データの増しのために Data Augmentation を行い、mixup などに加え、風や虫や雨音や IC レコーダーによる周波数フィルターを再現する処理を加えた。なお、クラスごとの教師データ数には大きなばらつきがあるため、分割はクラスごとに行うとともに、各クラスの学習する画像数が等しくなるように、画像の少ないクラスはアップサンプリング、多いクラスはダウンサンプリングを行った。

4. 評価用音源を用いた識別精度の確認

学習済みのモデルを用いて、2018 年 9 月 6 日~10 月 8 日の 17 時以降の夜間に沖縄で録音した延べ 572 時間の音源を識別した。なお、この音源は教師データの作成には使用していない。また、同じ音源を全て耳で聞いて確認し、1 鳴き 1 区間のヤンバルクイナが鳴いた正解区間を求めた。

このヤンバルクイナの識別結果と正解を比較した結果、図 2 に示す ROC 曲線から AUC が 0.83 となった。また、閾値ごとに求めた F 値の最大値は 0.48 であった。閾値を 0.5 とした場合の感度は 0.29 (451 / 1531)、適合率は 0.96 であった。

次に、誤識別の原因を確認するために、正解区間の一部でも抽出できているか否かで偽陰性と偽陽性の評価を行った。閾値 0.5 で偽陰性を示した 34 区間は合計で 483 秒間あり、そのスペクトログラムを確認すると、鳴き声の音圧が低いか、RR-XP007 による録音で 20 kHz 以上の音がカットされた上に音圧が低く多数の虫が同時に鳴いたものであった。ヤンバルクイナの教師画像は全て 22 kHz まで音圧が存在したため、学習データと異なる条件となり大きな尤度を得られなかったものと考えられる。なお、RR-XP007 の音源は 34 区間中 26 区間を占めていた。

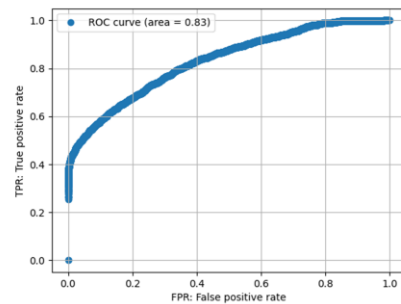


図 2 評価用音源に対する ROC 曲線

一方で、閾値 0.5 で偽陽性を示した 14 区間の合計 70 秒のスペクトログラムと音源を確認すると、うち 3 区間がヤンバルクイナの鳴き声であることを否定できず、残りの 11 区間は全てクチキコオロギとリュウキュウカネタタキとマダラコオロギが鳴いていた上に、雨音などが写り込んでいた。クチキコオロギは山原地域で頻繁に鳴く虫の一種であり、ヤンバルクイナの教師画像にもしばしば写っていた。

本研究ではクチキコオロギがヤンバルクイナの教師画像に写っていた場合は「ヤンバルクイナ」と「その他の虫」というラベルを付けて学習したが、「その他の虫」は 150 種類ほどの虫の鳴き声の集合であったので、画像パターンを十分に学習できなかったと考えられる。

なお、感度から予測されるほど偽陰性が悪くなかったことから、現時点では鳴き声が徐々に消えていく音圧の低い部分の検出ができていないものと考えられる。

5. おわりに

本研究は野鳥の環境調査において環境音から複数の野鳥を自動的に識別することを目的とし、IC レコーダーを用いて収集した音源からスペクトログラムを作成して VGG16 の転移学習を行った。2018 年に沖縄県で録音した音源を用いてヤンバルクイナの識別を行ったところ、AUC で 0.83 となった。未検出となる鳴き声は比較的少なく、既にヤンバルクイナの生息状況調査に使えるものと考えられる。今後はクチキコオロギの学習による偽陽性の削減と、異なる IC レコーダーを用いた場合の識別率の向上が求められる。

謝辞

本研究を遂行するにあたり、バードリサーチ様と多数のご協力者様、および東京大学サイバーフォレスト様に音源をご提供いただきました。この場を借りて、深く感謝申し上げます。なお、本研究は JSPS 科研費 21K14280 の助成を受けたものです。

参考文献

- [1] 太田 望, 延原 肇, 徳江義宏, 今村史子, 太田敬一, 夏川遼生, “深層学習による野外録音音声からの動物の鳴き声検出-オオタカを事例として-”, 情報処理学会第 80 回全国大会講演論文集, pp. 379-380, (2018).
- [2] 池長 裕史, 儀間 朝治, “ヤンバルクイナ(Rallus okinawae)の鳴き声とデュエットについて”, 山階鳥類研究所研究報告, Vol.25, No.1 (1993).