

奄美大島に生息する希少種の 鳴き声自動認識に関する研究

上村 優介¹ 福元 伸也¹ 鹿嶋 雅之¹ 渡邊 睦¹
榮村 奈緒子² 鷗川 信²

概要：2021年に奄美大島は「生物多様性」を理由に世界遺産に登録された。登録を維持するためには、生息する野鳥の観測を行い、希少種の存在を確認し続けなければならない。本研究では、野鳥を自動識別できるシステムの開発を行うことを目的とする。自然音から鳴き声を切り出し、スペクトログラム画像へ変換し、畳み込みニューラルネットワークで学習を行った。その際、マイクの風切り音を除去する処理を施し、識別精度の向上を図った。結果、風切り音を除去したことで識別精度が向上し、先行研究の手法であるハト類の2段階識別で最も高い精度で識別ができることが分かった。今後の課題として、自然音から野鳥の鳴き声を切り出す過程の自動化を目指す。

キーワード：分類学習、画像特徴抽出、画像分類

Research on automatic recognition of calls of rare species inhabiting Amami-Oshima Island

YUSUKE UEMURA¹ SHINYA FUKUMOTO¹
MASAYUKI KASHIMA¹ WATANABE MUTUMI¹
NAOKO EIMURA¹ SHIN UGAWA¹

Abstract: In 2021, Amami-Oshima Island was registered as a UNESCO World Heritage Site due to its 'biodiversity.' To maintain this registration, ongoing observation of the resident bird species is necessary to confirm the presence of endangered species. This study aims to develop a system capable of automatically identifying birds. It extracts vocalizations from natural sounds, converts them into spectrogram images, and trains a convolutional neural network. In this process, techniques for removing microphone wind noise were applied to improve identification accuracy. As a result, the removal of noise led to improved identification accuracy, allowing for the highest precision in the two-stage identification of pigeon species, as demonstrated in previous research methods. An upcoming challenge is to automate the process of extracting bird vocalizations from natural sounds.

Keywords: Classification learning, Image feature extraction, Image classification

1. 研究の背景と目的

2021年、奄美大島は「奄美大島、徳之島、沖縄島北部及び西表島」として世界自然遺産に登録された。登録の主な理由は、「生物多様性」とされている。これは、現地に生息する固有種や希少生物の存在が高く評価されていることの影響が大きいといえる。確かに奄美大島には、ルリカケスやアマミノオオウサギをはじめとする様々な固有生物が生息している。世界自然遺産への登録に伴い、登録を維持するためには、生息する生物の観測を行い、希少種の存在を確認し続けなければならない。そこで、警戒心が強いため近距離からの映像撮影が難しく、存在を確認するための人的負担が大きい「野鳥の観測」に目を付けた。野鳥は、距離がある場合は双眼鏡による目視での観測が行われるが、長時間飛行する鳥や夜行性の鳥に関しては鳴き声による観測が一般的である。

現在、録音機器の設置および音源の回収から、大量の音声データをもとにした野鳥の存在の確認に至るまで、ほぼすべての工程で人力による調査が行われている。数百時間の自然音の中から、数秒の鳴き声と特定する作業を、対象種の数だけ行わなければならない。本研究ではこのような作業負担を軽減し、自動で野鳥を識別できるシステムの開発を行うことを目的とする。

2. 外部状況

関連研究では、畳み込みニューラルネットワーク(CNN)で、前処理を施したスペクトログラム画像の分類を行った。結果、ランダムフォレスト、多層パーセプトロンが80%前後の精度であったのに対して、90%という高い精度で識別できた。ただし、対象は絶滅危惧種のサンバ1種のみである[1]。

また、先行研究では、スペクトログラム画像を用いた識

¹ 鹿児島大学 工学部
Faculty of Engineering, Kagoshima University, Japan
² 鹿児島大学 工学部
Faculty of Agriculture, Kagoshima University, Japan

別において、類似した特徴を持つ鳥を含む6種の野鳥の分類に成功している[2].

さらに、東谷らは、周波数帯域パワーを用いたニューラルネットワークによる識別を行っている[3]. 野鳥12種につき全体の識別率が94.29%と高い精度を出せているほか、60%を下回っているものがないため、周波数帯域パワーによる識別について有用性があるといえる.

本研究では、奄美大島に生息する10種についてスペクトログラム画像でCNNを用いて識別を行う. なお、スペクトログラム画像を用いるのは、時間的変化や周波数特性を画像特徴としてみることができるという性質を利用するためである. また、事前に風切り音を計測・除去し、スペクトログラム画像上での特徴を変化させたデータを用意し、比較実験も行う.

3. 提案手法

3.1 スペクトログラム画像の取得

本研究では、スペクトログラム画像による画像分類を行う. スペクトログラム画像とは、音声信号や周波数成分の時間変化を視覚化するために使用されるグラフィカルな表現方法である(図1). 横軸には時間を、縦軸には周波数を、色の濃淡によって周波数成分の強度を表している. 明るい色は高い強度を表し、暗い色は低い強度を表している.

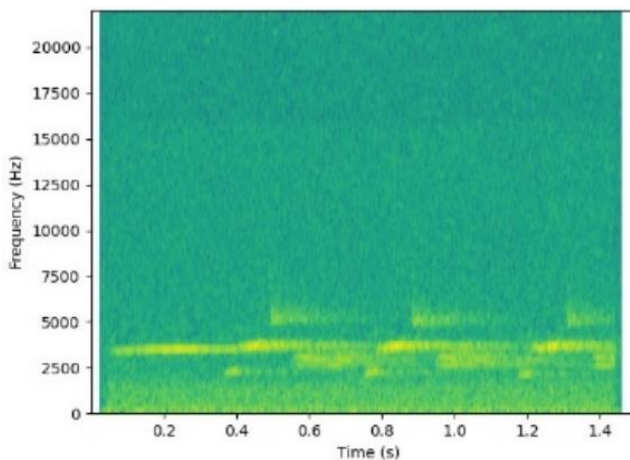


図1 アカヒゲのスペクトログラム画像
Figure 1 Spectrogram image of Redbeard

先行研究[4]より、奄美大島のオオトラツグミについて、スペクトログラム画像にて存在を確認しており、本画像を用いての識別方法は無難なものだと考える. 当研究より、鳥や動物における鳴き声の存在確認にはスペクトログラム画像は有効であるといえる.

本研究でも、識別対象の鳥の鳴き声についてスペクトログラム画像に出ている特徴をもとに画像識別を行う.

3.2 識別手法

本研究は以下(図2~4)の構成に沿って識別を行う. 手法1

は、風切り音を含む音源をもとにスペクトログラム画像を生成し学習に用いる. 手法2は、使用するデータを、風切り音を除去したスペクトログラム画像に変更する. 手法3では、ハト類を2段階で識別する.

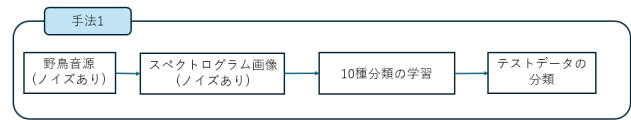


図2 野鳥識別の手法1

Figure 2 Methods of Bird Identification 1

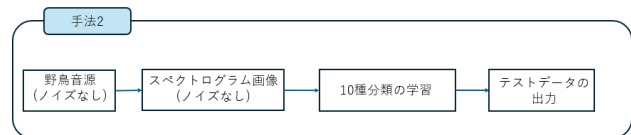


図3 野鳥識別の手法2

Figure 3 Methods of Bird Identification 2

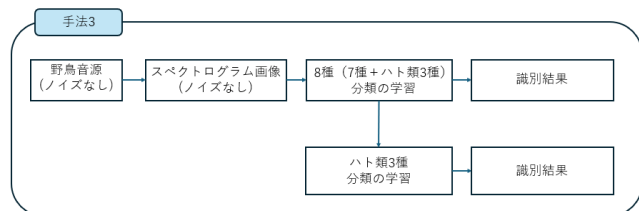


図4 野鳥識別の手法3

Figure 4 Methods of Bird Identification 3

機械学習を行うにあたって、全10種の鳴き声のデータは、各100枚のスペクトログラム画像を用意し、80枚を訓練データ、10枚を検証データ、10枚をテストデータに分割し、その順番を入れ替えることによって10パターン(図5)のモデルで実行する. 結果として、学習に使用していないデータの識別結果が100回分用意できることになり、その正答率を野鳥の識別精度として算出する.

パターン	train	validation	test
A	0~79	80~89	90~99
B	10~89	90~99	0~9
C	20~99	0~9	10~19
D	30~99,0~9	10~19	20~29
E	40~99,0~19	20~29	30~39
F	50~99,0~29	30~39	40~49
G	60~99,0~39	40~49	50~59
H	70~99,0~49	50~59	60~69
I	80~99,0~59	60~69	70~79
J	90~99,0~69	70~79	80~89

図5 10パターンの学習モデル

Figure 5 10 Patterns of learning models

3.3 畳み込みニューラルネットワーク

実験で用いる畳み込みニューラルネットワーク(図 6)について説明する。

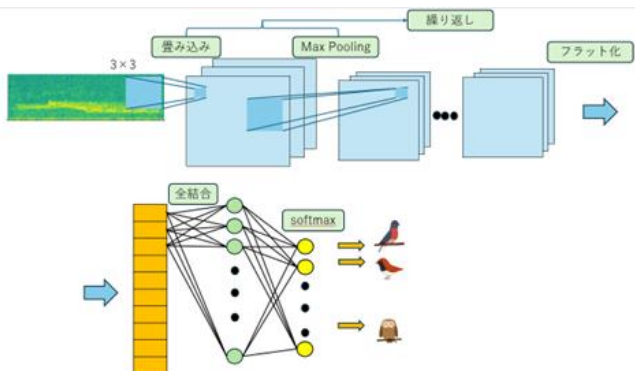


図 6 畳み込みニューラルネットワーク
Figure 6 Convolutional Neural Network (CNN)

まず、画像を入力し、畳み込みを行う。畳み込みでは Keras の Conv2D を使用し、 3×3 のフィルタを 64 枚使用している。そしてフィルタ後の画像に活性化関数「ReLU」を実施する。

次に、プーリング層を作成する。 2×2 の大きさの最大プーリング層を作成する。ここでは 2×2 の領域で最大の数値を出力するようにする。この最大プーリング層では単純にカーネルの中の最大値を特徴として抽出している。それに畳み込み演算を行う。上記同様、 3×3 のフィルタを 64 枚使用している。それに活性化関数を適応し、再度プーリングを行う。畳み込みとプーリングの終了後、画像データを列ベクトルにフラット化する。全結合では、出力を 256 にした。ソフトマックス関数を用いて、各出力の予想確率を計算させている。バッチサイズは 10 でエポック数は 20 である。

4. 実験

4.1 データの採取

本研究は、鹿児島大学農学部との共同研究である。データについては、鹿児島大学農学部の榮村先生が録音した現地の自然音と、そこから検出された各種野鳥が鳴いている時間などをまとめた Excel ファイル(図 7)をもとに研究を進めた。同一ファイルは、どの音声がどの鳥なのかというラベル付けがされている。自然音を含む音声ファイルを読み込み、鳴き声区間を切り出すプログラムで抽出した。

鳴き声を含む 自然音のファイル名	鳴き声 開始地点	鳴き声 持続時間	周波数 最小値	周波数 平均値	周波数 最大値	種別ID	鳥種名
IN FILE	OFFSET	DURAT	Fmin	Fmean	Fmax	ID	sp
220313_uk_one_66_m	136.225403	0.458277	473.73	564.648	689.063	aob	アオバズク
220626_uk_one_95_m	749.548767	0.404989	473.73	592.991	689.063	aob	アオバズク

図 7 野鳥の鳴き声に関する情報
Figure 7 Information regarding bird vocalizations

録音時間は、朝昼の 3 時間である。場所については、役勝唐川、役勝赤畑、役勝嘉徳又・長田、役勝丸畑、はつの、三太郎、金作原、嘉徳川内山、網野子亥ノ川である。役勝嘉徳又・長田、嘉徳川内山、網野子亥ノ川以外の箇所は、設置場所の地形を尾根と谷にそれぞれ分けた。

さらに、標高は 50~300m で観測している。野鳥によっては、あまり地上に降りてこない種や水辺の近くで生息する種がいる。観測区域が広域であることは、音声による観測において有利にはたらくと考えている。

録音対象は、奄美大島の固有種について、アオバズク・アカショウビン・アカヒゲ・オオトラツグミ・カラスバト・ズアカアオバト・リュウキュウキジバト・リュウキュウコノハズク・ルリカケスである。そのほかにも、キツツキのドラミングを録音している。これら 10 種が今回の識別対象である。

Excel ファイルで指定された音源数は、鳥種ごとにばらつきがあった(図 8)。しかし、学習する際にデータ量のばらつきがあつては、偏りが生じてしまうため望ましくない。そこで本研究では、最小音源数であるアカショウビンの 140 個と先行研究で使用したデータが 50 個であったことから、区切りの良い 100 個を使用することにした。その選び方は、先行研究で使用されていたデータを含むように、収録された音源の古い順から 100 音源とした。

野鳥	音源数
アオバズク	388
アカショウビン	140
アカヒゲ	283
オオトラツグミ	324
カラスバト	212
キツツキ(ドラミング)	271
ズアカアオバト	193
リュウキュウキジバト	346
リュウキュウコノハズク	481
ルリカケス	361

図 8 鳥種別の取得音源数
Figure 8 The number of audio recordings obtained per bird

また、Excel ファイルで指定されていない自然音区間から、無音状態の音源を 200 個取得した。これをもとに、処理すべき音響特徴について検討した[5]。無音状態では、20dB~50dB ほどの風切り音が含まれることがわかった。したがって、平均値である 30dB を閾値として、音量を下げる処理を施したデータも用意した(図 9)。また、識別対象の周波数帯域は 7kHz 以下に集中しているため、スペクトログラム画像のトリミングを行った。

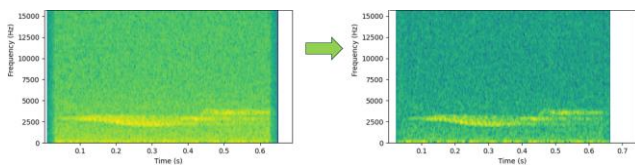


図 9 風切り音の除去によるスペクトログラム画像の変化
Figure 9 The change in spectrogram images due to wind-noise removal

4.2 風切り音を含む画像での 10 種分類

手法 1 では、Excel ファイルで指定された、野鳥が鳴いている箇所を切り出した音源を、スペクトログラム画像に変換し畳み込みニューラルネットワーク (CNN) で学習した。そのモデルをもとに、学習に用いていないテストデータによってテストを行い、その出力を確認する。使用する画像は以下のようなものである(図 10)。

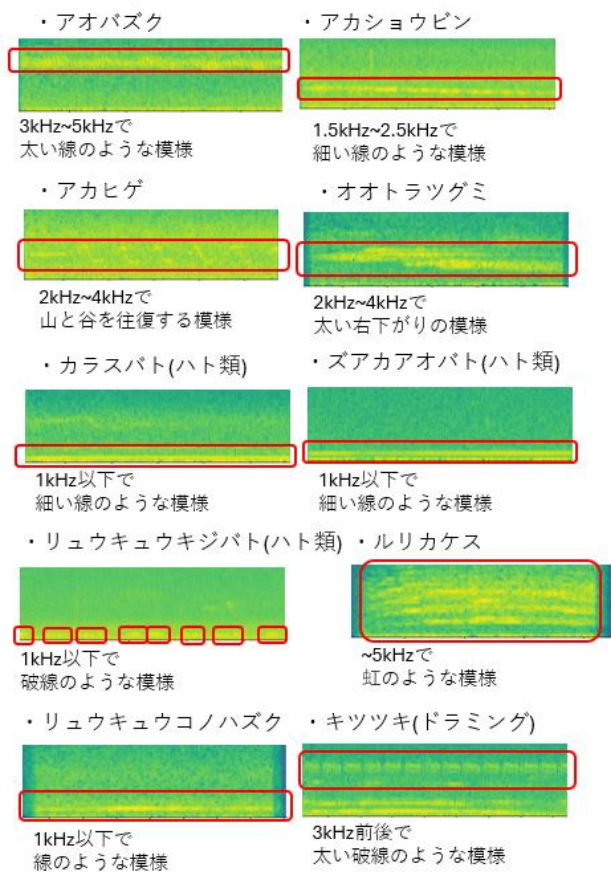


図 10 対象種のスペクトログラム画像

Figure 10 Spectrogram images of the target species

4.3 風切り音を排した画像の 10 種分類

手法 2 では、前述の図 9 に示した、風切り音を除去したスペクトログラム画像を用いて 10 種分類の学習を行う。学習は、手法 1 と同様に、図 5 に示した 10 パターンで行う。これは、手法 1 との対照実験であり、風切り音の除去の前後で識別結果がどのように変化するかを確かめるものである。

4.4 2 クラス分類を利用した分類

手法 3 では、カラスバト・ズアカアオバト・リュウキュウキジバトをハト類として検出し、その後ハト類同士を識別するという、先行研究で用いられた手法について学習を試みる。これは、ノイズの除去によって、特徴が似ている鳥種の分類においても精度の向上が期待できるだろうという仮定を検証するものである。つまり、7 種+ハト類の 8 種分類(手法 3-1)とハト類 3 種分類(手法 3-2)を行う。手法 3-1 は、ハト類に特化する分類なので、さらに周波数帯域を 1kHz 以下に絞ったスペクトログラム画像を利用して学習を行う。以下(図 11)に、使用するスペクトログラム画像の例を示す。

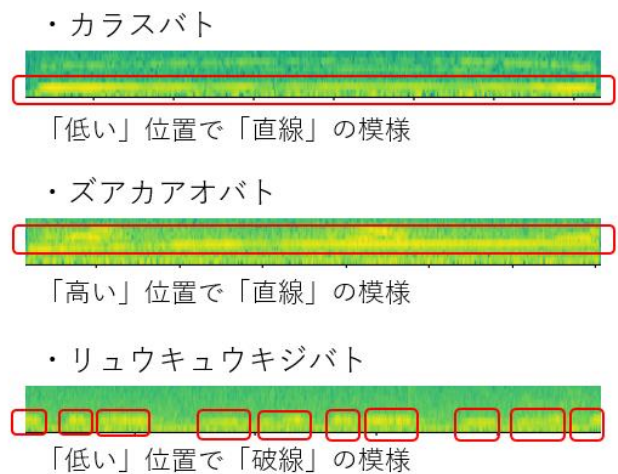


図 11 手法 3-2 で用いるスペクトログラム画像
Figure 11 The spectrogram images used in Experiment 3-2

5. 実験結果

5.1 手法 1 の結果

手法 1 については、風切り音を含むスペクトログラム画像を用いて、対象種 10 種を分類している。以下の表 1 は手法 1 についての識別結果を示している。

表 1 手法 1 の結果

Table 1 Result of Experiment 1

野鳥の名称	種別識別率(%)
アオバズク	93
アカショウビン	97
アカヒゲ	93
オオトラツグミ	92
カラスバト	39
キツツキ(ドラミング)	78
ズアカアオバト	90
リュウキュウキジバト	93
リュウキュウコノハズク	88
ルリカケス	98

表 1 より、ほぼ全種が 85%前後で識別できていることが分かる。最も識別率が高かったのは、ルリカケスの 98%である。次いでアカショウビンの 97%、アオバズク、アカヒゲ、リュウキュウキジバトの 93%となっている。しかし、中にはキツツキ(ドラミング)の 78%やカラスバトの 39%というように精度が低い種もあることがわかった。

5.2 手法 2 の結果

手法 2 については、風切り音を除去したスペクトログラム画像を用いて、対象種 10 種を分類している。以下の表 2 は手法 2 についての識別結果を示している。

表 2 手法 2 の結果

Table 2 Result of Experiment 2

野鳥の名称	種別識別率(%)
アオバズク	91
アカショウビン	97
アカヒゲ	85
オオトラツグミ	93
カラスバト	88
キツツキ(ドラミング)	65
ズアカアオバト	88
リュウキュウキジバト	92
リュウキュウコノハズク	97
ルリカケス	97

表 2 より、ほぼ全種 90%前後で識別できていることがわかる。アカショウビン、リュウキュウコノハズク、ルリカケスは 97%の識別精度であった。カラスバトに関しては、ノイズを除去した前後で 39%から 88%へと 49%も識別精度が向上している。また、最も識別率が悪かったのは、キツツキ(ドラミング)の 65%である。

5.3 手法 3-1 の結果

手法 3-1 は、特徴が似ているカラスバト・ズアカアオバト・リュウキュウキジバトをハト類として 1 括りで分類する画像識別である。つまり、8 種の分類である。ここでも手法 2 と同様に、風切り音を除去したスペクトログラム画像を用いる。以下の表 3 は、手法 3-1 の識別結果を示す。

表 3 手法 3-1 の結果

Table 3 Result of Experiment 3-1

野鳥の名称	種別識別率(%)
アオバズク	88
アカショウビン	97
アカヒゲ	90
オオトラツグミ	97
キツツキ(ドラミング)	88
リュウキュウコノハズク	98
ルリカケス	98
ハト類	99

表 3 より、ほぼ全種 95%前後で識別できていることがわかる。1 括りで分類したハト類について最も精度が高く、99%となった。手法 1、手法 2 で識別精度が低かったキツツキ(ドラミング)に関しては、88%まで精度が向上した。最も悪い識別精度で、キツツキ(ドラミング)とアオバズクの 88%であった。

5.4 手法 3-2 の結果

手法 3-2 は、カラスバト・ズアカアオバト・リュウキュウキジバトの 3 種を分類している。以下の表 4 は手法 3-2 についての識別結果を示している。

表 4 手法 3-2 の結果

Table 4 Result of Experiment 3-2

野鳥の名称	種別識別率(%)
カラスバト	94
ズアカアオバト	100
リュウキュウキジバト	98

表 4 から、ほぼ 97%前後で識別できていることがわかる。ズアカアオバトに関しては、誤識別が起こることはなく 100%で分類している。また、カラスバト-リュウキュウキジバト間で互いに誤識別を起こしていた。

6. 考察

手法 1 の結果から、ルリカケスやアカショウビンのように、他の鳥とスペクトログラム画像上での特徴が似ていない鳥に関しては、風切り音を含むスペクトログラム画像による畳み込みニューラルネットワークでも識別精度は高いことがわかる。同時に、カラスバトとキツツキ(ドラミング)のように、スペクトログラム画像上での特徴が似ている、あるいは鳴き声の重複(図 12)が起こっている場合では、識別精度が低くなることがわかる。

今回対象とした風切り音以外にも、雨音や小動物の足音など、鳴き声以外の特徴が音源に含まれていることがある。誤識別したテスト画像を確認したところ、その大半が鳴き声の重複によるものであることがわかった。次いで、風切り音以外の不適切な特徴(雨音や小動物の足音など)が現れているものであった。これは識別において大きな障害となりうるため、様々な雑音に対する専用の除去処理を用意する、あるいは該当音源を分離する処理[6]が必要になるだろう。

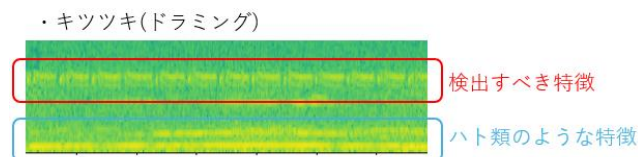


図 12 誤った分類を起こした事例
Figure 12 Example of misclassification

手法2の結果から、主に手法1で識別精度が低かった種について、識別精度の底上げをすることができた。しかし、カラスバトと相互に誤識別を引き起こしていたキツツキ(ドラミング)は、78%から65%まで精度が悪化している。これは、風切り音の処理によってキツツキ(ドラミング)として識別するための特徴量が失われた可能性が考えられる(図13)。キツツキのドラミングは鳴き声ではないので、画像特徴よりも音色特徴によって識別するほうが適しているだろう。

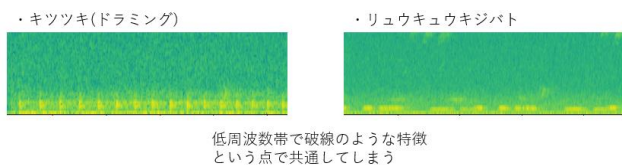


図13 風切り音の除去で出た悪影響

Figure 13 The adverse effects of removing slight wind noise

手法3では、風切り音を除去したスペクトログラム画像を用いた学習を、先行研究で開発された2段階での分類手法で実施した。ハト類として99%識別でき、ハト類3種も高い精度で識別できることがわかった。特徴が似ている鳥をまとめて検出したことで、精度が低かったキツツキ(ドラミング)、カラスバトに関しては精度が大幅に向上した。

手法2と手法3を比較すると、手法3のほうが精度が高い。これは、鳴き声の重複による誤識別を最も引き起こしていたハト類がほかの鳥の識別に与える影響が小さくなったと考えることができるだろう。ただし、アオバズクのように識別精度が低くなった鳥種もあった。その誤識別の相手はハト類が多かったことから、1kHz以下におけるスペクトログラム画像の特徴量が野鳥識別全体に与える影響が大きいといえるだろう。その裏付けに、手法3-2の精度が全体で最も高かった。

7. まとめと今後の課題

本研究で対象としていた10種の野鳥について、高い精度で識別できていることがわかった。

そして、本研究では先行研究の6種識別を10種識別に拡張すること、そして音声に成形処理を施し識別精度を向上させることを目標に置いていた。実験では、風切り音を含むスペクトログラム画像を用いた分類と風切り音を除去したスペクトログラム画像による分類の対照実験および類似特徴を識別するための2段階モデルによる分類を行った。その結果、誤識別の原因のうち、大半が鳴き声の重複と想定外のノイズ特徴によるものであるとわかった。適切な除去・音源分離の方法を開発することでさらなる精度の向上が期待できるだろう。

また、本研究では使用したデータが100データであり、

テストデータを除くと、訓練データと検証データに関しては90枚であった。先行研究からデータ量は2倍に増加させたが、今後も継続的なデータの追加を行い、適宜、識別実験を行う必要がある。

今後の取り組みとしては、第一に、自然音から自動で野鳥の鳴き声音源を切り出す処理を実装することである。本研究において最も時間がかかった作業工程であり、システム全体の実行時間に最も影響が出ると思われるからだ。

第二に、自動で切り出された音声による分類精度の検証である。この工程では、自動で切り出された音声の中で正しく鳥が鳴いているか、対象種以外の鳴き声や関係のない音声だった場合、野鳥の分類から排除できるかが重要視される。

第三に、システム全体の自動化に向けたサーバーの設置およびプログラムの改良である。入力音源に強風や雨音が入っているかという事前情報をクライアントから受け取ることで、その特徴を排除するような処理を施すというようにできれば、分類の前処理工程にかかる実行時間の削減が行えるだろう。

参考文献

- [1] 前川侑子, 田口華麗, 牛込祐司, 佐藤匠, 小林啓悟, 芳賀智宏, 町村尚, 東海明宏, 松井孝典, "AI技術による鳥類の鳴き声モニタリング手法の検討~サシバを事例として~", *Bird Research* vol.18 p71~86, 2022.
- [2] 眞島京音, 福元伸也, 鹿嶋雅之, 渡辺睦, 鶴川信, 榮村奈緒子, "奄美大島に生息する希少種の鳴き声自動認識に関する研究", 火の国シンポジウム, A8-1, 2023/03/14
- [3] 東谷幸治, 三田長久, 牧野洋平, "音声情報によるニューラルネットワークを用いた夜行性野鳥の識別", *FIT2006*(第5回情報科学技術フォーラム) p355-356
- [4] NPO法人奄美野鳥の会, 鳥飼久裕, 水田拓, 島隆徳, 鈴木真理子, "世界自然遺産登録地奄美大島における録音装置を用いた鳥類モニタリング手法の開発と次世代の調査従事者の育成" 自然保護助成基金助成成果報告書 vol.31(2022)p193-201
- [5] mp3の加工について, <https://271108.hatenablog.com/entry/2022/02/27/191217> (参照 2023-12-18)
- [6] 深層学習を音源分離に応用, <https://271108.hatenablog.com/entry/2022/02/27/191217> (参照 2024-01-11)
- [7] 川原竣介, 平川凜, 中藤良久, "音声スペクトルの時間変化協調による明瞭正改善手法の提案", *産業応用工学会論文誌* vol.6 pp51-54(2018)
- [8] スマホやPCで鳴き声をAIで解析して鳥の種類を判別できる「BirdNET」が登場, <https://gigazine.net/news/20210802-birdnet/> (参照 2023-09-07)
- [9] 鳴き声から野鳥を識別するアプリ Song Sleuth (ソングスルース), <https://chizaizukan.com/property/337/> (参照 2023-11-13)
- [10] 「音声情報を活用した生物の生息調査システムの開発」が2018年日経地球環境技術賞(第28回)において優秀賞を受賞, <https://pr.fujitsu.com/jp/news/2018/10/31-1.html> (参照 2023-12-20)