

GO TO COOT

富田 溪斗 (筑波大学 生物学類) 指導教員: 徳永 幸彦 (筑波大学 生命環境系)

【背景及び目的】

茨城県の蓮根生産量は日本一であり、その大半は霞ヶ浦周辺で生産されている。同時に、霞ヶ浦はオオバンやカモ類など水鳥の生息地で、しばしば隣接する蓮田に侵入し、蓮根を食害しているとみられており、年間の被害額は約2億円になると試算されている。これら害鳥の食害を防ぐ目的で土浦入沿岸の一部の蓮田には防鳥ネットが仕掛けられているが、サンカノゴイやコウノトリなどの絶滅危惧種が羅網死するという事故も起きている。こうした中、オオバンやカモ類などの採餌行動は明らかになっておらず、さらなる調査が求められているが、夜間の目視による蓮田での採餌行動観察は難しい。目視によらない観察方法を確立することで、新たな生体情報を得、それを防鳥ネットに頼らない防除方法に活かしていくことが希少種保全および農業活動の両方において重要な課題となっている。

近年では、鳥類の鳴き声を自動記録装置などで記録し、得られた音声データをニューラルネットワークなどの自動認識モデルで認識し、生態情報として活用する試みが盛んになっている。鳥類における特定種の鳴き声の認識モデルの活用法としては、(1)在情報の自動取得、(2)各時刻における鳥類の活性、(3)位置推定および個体数推定などが挙げられる。本研究では、オオバンの鳴き声を認識するモデルを畳込みニューラルネットワーク(以下、CNN)により作成した。

【方法】

(1) 教師音源データの収集と前処理

オオバンの鳴き声のサンプリングは、茨城県土浦市霞ヶ浦周辺で2020年6月7日、同月21日の16時~19時の間に行った。音源は、48kHz・16bitのmp3形式にし、Audacityを使用し、鳴き声を切り取り、計181個の鳴き声データを得た。また、鳴き声以外の環境音を含む音源を切り取り計210個得た。得られた音源データを、Python音処理パッケージのlibrosaによってSTFT、画像化し、幅32px、高さ128pxの縦長の教師画像データとした。このうち、8割を教師データ、2割を評価用データとした。

(2) モデルの作成

CNNの実装には、Pythonの機械学習パッケージであるPytorchを用いた。ニューラルネットワークの構造は、図1のようになっている。

(3) 評価実験

土浦市で収集した音源で訓練したモデルの性能を評価するため、鳥の鳴き声データベースであるxeno-cantoから、日本のオオバン(*fulica atra*)の鳴き声データを集め、訓練したモデルで連続音声認識をおこなった。収集したオオバン鳴き声データは全13音源、172秒であった、各音源を4096データ点でシフトしながら、STFTにより32px×128pxの画像にし、1枚ずつ訓練済みCNNモデルに入力して認識結果を得た。認識結果は図1のように陽性である部分を網掛けしている。認識結果の図1から、真陽性、偽陽性、偽陰性の数を手作業で集計し、性能評価指標として精度と再現率を算出した。

精度及び再現率の算出式を以下に示す。

$$\text{精度} = \frac{\text{真陽性}}{\text{真陽性} + \text{偽陽性}}$$

$$\text{再現率} = \frac{\text{真陽性}}{\text{真陽性} + \text{偽陰性}}$$

精度は陽性と認識したものが実際に陽性であったものの割合であり、再現率は実際に陽性であるものの内、陽性と認識されたものの割合である。

【結果および考察】

評価実験の結果、真陽性が148回、偽陰性は19回、偽陽性は1回であり、精度は0.993、再現率は0.886であった(表2)。この精度により、今回作成したモデルが陽性であると判断したものは99%がオオバンの鳴き声であり、また再現率から、89%の鳴き声を漏らさずに認識できると推測できる。

データベース上の音源では高い精度及び再現率が出たものの、ノイズなどを含まないきれいな音源をxeno-cantoにユーザーがアップロードしている可能性があり、その条件では精度や再現率が高くなる可能性がある。実際の自然環境においては、風や車などが録音環境を濁す場合が多く、望ましいクリアな条件で録音ができることも限らないため、雑音を多く含む音源に対しても認識できるかどうか実用上の課題である。そして本研究の一つの活用法である、認識モデルを用いた位置推定について、これを行う際には個体の追尾が可能な頻度で、認識モデルが認識できる音条件で個体が鳴き声を発するかどうかという点についても今後実地での野外調査が必要になる。

表1 作成したニューラルネットワーク

Input Image	[128, 32]
Convolutional	[1,32,3]
MaxPooling	[2]
Batch Normalization	[32]
Drop Out	[64]
Convolutional	[32,64,3]
MaxPooling	[2]
Batch Normalization	[32]
Drop Out	[64]
Linear	[64,200]
Batch Normalization	[200]
Drop Out	[0.25]
Linear	[200,2]

表2 評価実験の結果

真陽性	148
偽陰性	19
偽陽性	1
精度	約0.993
再現率	約0.886

図1 認識結果(例)

