MC2　物体検出アルゴリズムを用いた日本の鳥類の鳴き声検出と種判別

Bird Song Detection and Species Identification using Object Detection Algorithm

指導教官　町村尚 准教授・地球循環共生工学領域

28H17066　堀尾亮太（Ryota HORIO）

Abstract: Monitoring bird species composition and diversity is quite effective to estimate environmental situation. And acoustic detection of bird songs can be used to establish automated monitoring system of multiple bird species. I developed a bird species identifer by recognizing vocalization spectrogram using object detection algorithms based on deep learning technology. In this study, object detection algorithms were fine-tuned ResNet-18 (SSD) and Darknet-53 (YOLOv3) pretrained with the imagenet to classify 14 bird species. I also applied data augmentation techniques to avoid overfitting and further improve model's generalized performance. SSD and YOLOv3 with bounding box adjustment and data augmentation were suitable for real time inference and high classification performance, respectively.

Keywords: Spectrogram, YOLOv3, SSD, Bird sound, Bird detection

１．背景

現状の環境状態や生物多様性を評価するための環境モニタリングに世界の関心が高まっている [1]．その中で，環境の変化に敏感な鳥を対象に，視界が不明瞭な夜間や木々が密集する森林内でも観察可能な音声モニタリングで種構成や多様性を把握することが期待されている [2]．代表的な先行研究では，畳み込みニューラルネットワークのマルチチャネル技術と転移学習などの画像認識タスクで鳥の鳴き声が予測されているが [3]，複数種の同時検出やリアルタイム識別に重要な FPS (Frames per Second) が考慮されていない課題がある．そこで本研究では，鳴き声のスペクトログラムから物体検出アルゴリズムを用いて高精度かつ高速に鳥の種類を検出できる識別器を開発することを目的とする．

２．分析方法

２．１　データベースとモデルの構築

図 1 に開発の全体像を示す．2007～2018 年に日本の 15 県で録音した 14 種類の鳥の種類のラベル付き音声ファイル 22,901 を収集した (wave 形式, サンプリング周波数 44.1 kHz, 16 bit)．各音声ファイルを 100 ms 間隔でサンプリングし, 3 dB バースト点を検出した．バースト点を中心に前後 500 ms, 周波数帯 0.1～12 kHz の FFT スペクトログラムを生成した．5,049 の FFT スペクトログラムに対して 10,672 の鳴き声の位置を示す Bounding Box と鳥の種類をアノテーションした．この際，鳴き声の周波数情報の付与するために Bounding Box を画像サイズに調整する処理 (BB 処理) を行い，9:1 で学習用とテスト用のデータに分割した．学習用データには 5 つの Data Augmentation (Horizontal transition, Cutout, Random erase, Salt and pepper, Mixup) を行った．ResNet-50 をベースにした SSD: Single Shot MultiBox Detector (512×512 pix) と Darknet-53 をベースにした YOLOv3: You Look Only Once v3 (416×416 pix) を 6 種類構築した (**表 1**).

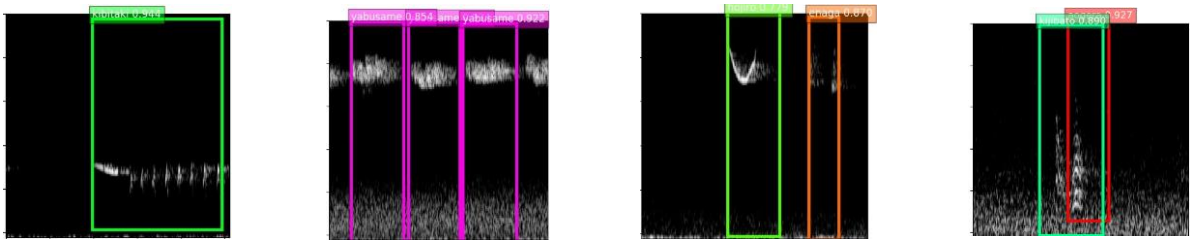
図 1　開発の全体像

２．２　モデル評価

識別精度には，全体精度の平均平均適合率 mAP (mean Average Precision)，IoU (Intersection over Union) 0.5，0.75 をそれぞれ閾値とした mAP@.5，mAP@.75，リアルタイム処理速度を表す FPS で評価した．

表 1　各モデルの詳細

No.	Method	バックボーン	BB 処理	Data Augmentation
1	SSD	ResNet-18		
2	SSD	ResNet-18	✔	
3	SSD	ResNet-18	✔	✔
4	YOLOv3	Darknet-53		
5	YOLOv3	Darknet-53	✔	
6	YOLOv3	Darknet-53	✔	✔



(a) キビタキ検出 (単一種)　(b) ヤブサメ検出 (単一種)　(c) ホオジロ・エナガ(複数種)　(d) アオグラ・キジバト(重複)

図 2　SSD モデル No.3 を用いた検出例 (縦幅:周波数 0.1~12 [kHz], 横幅:時間 1 [s])

３．結果と考察

３．１　複数種の検出

図 2 に SSD モデル No.3 を用いた鳴き声検出のスペクトログラムの例を示す．**図 2(a)** では単一種かつ単一の鳴き声のキビタキ，**図 2(b)** では単一種かつ複数の鳴き声のヤブサメの検出例を示す．単一種では鳴き声数に関わらず十分な検出ができた．また**図 2(c)** では複数種かつ複数の鳴き声のホオジロ・エナガ，**図 2(d)** では複数種かつ複数の重複した鳴き声のアオゲラとキジバトの検出例を示す．これらより，複数の鳥が混在する状況下でも概ね鳥の鳴き声が分離して検出されることが示された．また**図 2(b)** と **図 2(d)** では低音域にノイズが生じているが，ノイズ環境下でも複数の鳴き声を検出できていることが示された．

表 2　各モデルのベンチマーク

No.	Method	mAP	mAP@.5	mAP@.75	Best Epoch	FPS
1	SSD	43.8	82.2	43.4	240	44.1
2	SSD	65.0	89.1	76.0	176	**49.3**
3	SSD	66.6	88.9	**77.8**	131	45.8
4	YOLOv3	42.3	85.1	34.3	182	21.4
5	YOLOv3	57.2	87.8	64.8	238	21.9
6	YOLOv3	**67.0**	**90.0**	77.6	**52**	21.1

※　本研究の mAP は IoU の 0.5 から 0.95 まで(ステップ 0.05 ずつ)の異なる mAP の平均値である mAP@[.5, .95]を表す.

３．２　各モデルのベンチマーク

表 2 に各モデルのベンチマークを示す．SSD モデル No.1~3，YOLOv3 モデル No.4~6 を比較した場合，mAP では概ね SSD モデルと YOLOv3 モデルには差がなく，BB 処理と Data Augmentation による精度の向上が見られた．AP と mAP@.5 では YOLOv3 モデル No.6 が 67.0%と 90.0%を達成した．一方で mAP@.75 では SSD モデル No.3 で 77.8%を達成した．Best Epoch では YOLOv3 モデル No.6 が最も早く学習が収束したが, FPS では SSD モデル No.1~3 が 40 fps 以上の高い値を示した．この結果から，速度を重視する利用環境では SSD モデル No.2，鳴き声の検出力を重視する場合には YOLOv3 モデル No.6 が良好な性能を発揮することが示唆された．

４．今後の課題

モデルの精度向上のためには，TridentNet などの新規の物体検出アルゴリズムの適用や深いバックボーン (ResNet-101) を使用することが挙げられる. また単一種や複数種の検出精度を評価するために, 新しい環境下や他の動物種の鳴き音を含んだ音源を利用して汎化性能を評価することが求められる. さらに, 実環境下でリアルタイム識別に適用をめざす場合, 識別モデルの FPS の向上を目指す一方で, 無線通信機器の通信速度などにもバランスよく考慮することが重要であると思われる.

参考文献

1)　松井孝典：研究事例: 深層学習技術による環境音識別，日本音響学会誌, 74, 7, 2018.

2)　Priyadarshani, N. et al.: Automated birdsong recognition in complex acoustic environments.; a review. Journal of Avian Biology, 49, 5, 1-27, 2018.

3)　Jiang-jian, X. et al.: Audio-only Bird Species Automated Identification Method with Limited Training Data Based on Multi-Channel Deep Convolutional Neural Networks, Reprentation Fusion, Information Sciences, 348, 209-226, 2018.