

テーマ番号	1EP43			
プロジェクト テーマ	和文	深層学習に基づく野生動物認識システムに関する研究		指導教員 長田 茂美 教授
	英文	A study on wildlife recognition system based on deep learning		
プロジェクト メンバー	4EP1-08 大久保 裕弥 (Yuya Okubo)		4EP4-23 古松 竜一 (Ryuichi Komatsu)	

Abstract In recent years, crop damage by wildlife has become a serious social problem in Japan. To resolve this problem, camera trap images are used to grasp habitat information of wildlife. However, as images increase, it becomes harder for researchers to examine huge number of images and videos. In this study, we developed methods for estimating the numbers of wildlife and recognizing essential information such as their behavioral patterns from images and videos using deep learning technology. Our methods realized 79% accuracy for numbers of wildlife estimation and 80% accuracy for behavioral pattern recognition.

Keywords wildlife, deep learning, numbers of heads estimation, instance segmentation, action recognition.

1. まえがき

野生動物による農作物被害は毎年約 200 億円にのぼり、深刻な社会問題となっている。この対策を講じるためには、野生動物の種類や頭数、行動等の生息情報の調査が必要であり、カメラトラップ（定点自動撮影カメラ）画像の活用で、これらの情報収集が効率化されている。しかし、調査者が大量の画像を目視で確認し、動物が写っている画像を見つけ、生息情報を読み取る労力の大きさが問題となっていた。この問題を解決するため、長田研究室では、画像から野生動物の種類や写っている領域を生息情報として認識し、記録する、深層学習に基づく野生動物認識システムを開発してきた。今年度は、新たな機能として、野生動物の頭数推定と行動認識の機能を開発し、評価実験により有用性を確認した。

2. 野生動物認識システム

図 1 に、野生動物認識システムの概要を示す。本システムは、昨年度までに開発した領域検出、種認識の機能に加え、本研究で開発した頭数推定、行動認識の機能を有する。以降、頭数推定と行動認識の機能とその実現手法について述べる。

2.1 頭数推定

カメラトラップ画像に写っている野生動物の頭数を推定する機能を開発した。本機能は、画像中の野生動物の領域検出とその領域の数え上げの処理から成る。頭数推定のための領域検出では、異種の物体だけでなく、同種の物体が重なって写っていても分割して個々に検出可能なインスタンスセグメンテーションを実現する必要がある。そこで、本研究では、深層学習に基づくインスタンスセグメンテーションの代表的な手法である Mask R-CNN[1]を用い

ることとした。Mask R-CNN では、大量の学習データが必要であるが、画像に野生動物領域の真値を付与する労力は大きく、大量の学習データ作成は困難である。このため、既存の大規模データセットである MS COCO[2]を用いることとした。MS COCO には、人や車など多種多様な学習データが含まれており、対象物の種類に依存しない共通の画像特徴を学習することができる。しかし、本研究が対象とする森林環境中の野生動物の画像はほとんど含まれておらず、野生動物画像のデータを少数でも用意し、追加で学習する必要がある。そこで、本研究では、野生動物画像を用いたファインチューニングを実施した。図 2 に、本手法による領域検出の結果を示す（検出した領域は白色の輪郭線の内部）。

2.2 行動認識

本研究における行動認識は、野生動物の「歩く」、「止まる」、「食べる」など生息情報として有用な行動パターンを映像から認識する機能である。この機能を実現するため、まず、野生動物の動きの特徴に関わる関節座標を抽出し、その時間的な変位から動きの特徴を捉えて行動パターンを分類する手法を開発した。

2.2.1 関節座標抽出

関節座標とは、人間や野生動物の関節や主要な体の部位について、映像中の各フレーム画像上での位置を示す座標である。本研究では、野生動物の映像を観察し、次の 11 箇所の関節座標を用いることとした。

対象関節：頭、首、右肩、左肩、右肘、左肘、腰、
右股、左股、右膝、左膝

関節座標の抽出には、深層学習に基づく関節座標抽出ツールである DeepLabCut[3]を用いた。DeepLabCut では、映像中の一部のフレーム画像の関節座標を人手で真値として

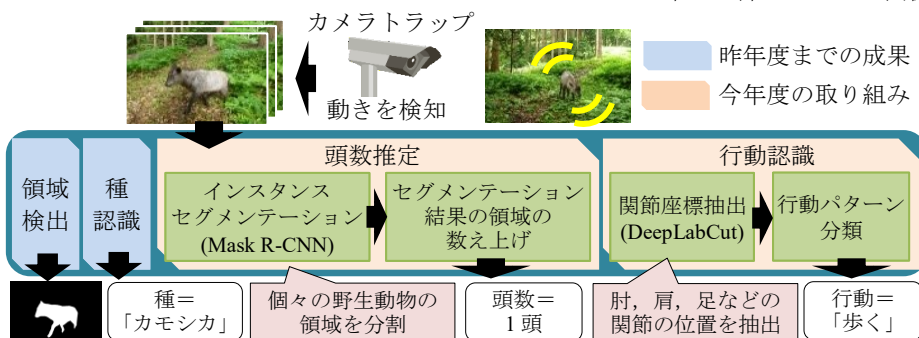


図1 野生動物認識システムの概要

Fig. 1 Overview of the wildlife recognition system.



図2 領域検出の結果

Fig. 2 Result of segmentation.

付与し、教師データとして学習することで、その他のフレーム画像の関節座標を抽出できる。本研究では、DeepLabCutを用い、対象関節の真値を著者が設定し、野生動物映像の関節座標を抽出した。図3は抽出結果を丸印で示した例である。DeepLabCutによる関節座標の出力データは、各フレーム画像上の11箇所の対象関節座標(x, y)を行として並べた22次元ベクトルを、フレーム画像の時間順に列として並べた行列データである。



図3 関節座標抽出の結果

Fig. 3 Result of joint coordinate extraction.

2.2.2 行動パターン分類

DeepLabCutで得られた関節座標データを用いて、行動パターンを分類する。本研究では、行動パターンとして「歩く」、「止まる」、「食べる」を対象とした。行動パターンを認識する手法としては、パターンマッチングを用いた。まず、映像から、「歩く」、「止まる」、「食べる」に該当するシーン（同じ動作が続くフレーム画像列）の一部を各行動パターンのモデルとして人手で選択する。次に、選択したシーンから各行動パターンの特徴を表す特徴ベクトルを計算する。特徴ベクトルの計算では、まず、シーン中で連続するフレーム a とフレーム b の間で各関節座標(x, y)の差分の絶対値をとった値を要素とするベクトル V_{ab} を求める。次に、シーン中で連続するフレーム画像ごとに得られるベクトル V_{ab} の平均ベクトルをシーンの特徴ベクトルとする。また、行動パターンのモデルとして複数のシーンを選択した場合は、各シーンについて求めた特徴ベクトルの平均ベクトルを行動パターンのモデルの特徴ベクトルとする。このように用意した各行動パターンのモデルの特徴ベクトルに対し、認識対象のシーンからも同様の方法で特徴ベクトルを求める。この認識対象シーンの特徴ベクトルに対し、各行動パターンのモデルの特徴ベクトルとのユークリッド距離を計算し、最も距離が小さい行動パターンを認識結果とする。

3. 評価

3.1 頭数推定の評価

頭数推定の評価では、石川県農林総合研究センター林業試験場に設置されたカメラトラップの画像を用いた。学習データは、2026枚の画像と各画像上の野生動物の領域（画素）を真値として人手で付与したデータである。テストデータとしては7725枚の画像を用いた。

表1に評価結果の推定頭数の正解率を示す。本評価では、MS COCOで学習したMask R-CNNを用いる「ファインチューニング前」の正解率と野生動物画像でファインチューニングを行った「ファインチューニング後」の正解率を求めた。また、頭数の違いによる差も分析するため、頭数別の正解率も求めた。表1に示すように、ファインチューニング前の正解率54%に対して、ファインチューニング後の正解率は79%であり、本手法の有用性を示すことができた。

一方、頭数別の評価では、頭数が増えるほど正解率が低下していた。この原因として、多重の重なりがある場合に誤認識していることがわかった。多重の重なりがある学習データはまだ少数であるため、学習データの拡充で解決を図ることが考えられる。

表1 頭数推定の正解率

Table 1 Accuracy of numbers of heads estimation.

	ファイン チューニング前	ファイン チューニング後
1頭	59%	81%
2頭	27%	67%
3頭	14%	52%
4頭以上	0%	52%
全体	54%	79%

3.2 行動認識の評価

行動認識の評価では、石川県農林総合研究センター林業試験場のカメラトラップで撮影したカモシカの映像を用いた。関節座標抽出の評価では、著者が目視で判定した真値の座標と抽出結果の座標とのズレの画素数を調べた。この結果、ズレは平均7ピクセルであり、この程度のズレであれば、野生動物の関節の動きの大きさ（フレーム画像間の変位の大きさ）に対して十分小さく、行動パターン分類に影響がない範囲であることが確認できた。

行動パターン分類の評価では、著者が行動パターンを判定したモデル用の8シーン、テスト用の5シーンを用いた。各行動パターンのシーン数を以下に示す。

- モデル：歩く(5), 食べる(2), 止まる(1)
- テスト：歩く(3), 食べる(1), 止まる(1)

評価の結果、誤認識は「歩く」を「食べる」とした1シーンのみであり、正解率は80%であった。誤認識の原因は、該当シーンの関節座標の動きを観察したところ、関節の部位により行動パターンとの関係の強さが異なるが、それが考慮されていない点であることがわかった。また、頭の縦の動きの反復を捉えるなどより高次の動きの特徴抽出が必要であることもわかった。こうした行動パターンを特徴付ける要因を人手で洗い出すことは困難である。このため、より大きなデータセットを整備し、機械学習で見出すアプローチが有用と考える。

4. むすび

本研究では、カメラトラップの画像や映像からの野生動物の頭数推定、および、行動認識の手法を開発し、有用性を評価した。今後は、さらなる精度向上を目指し、データセットの拡充や学習手法の改良でより実用的なシステムへと発展させていきたい。最後に、評価データをご提供頂いた石川県立大学の北村俊平准教授、石川県農林総合研究センターの関係各位に感謝致します。

参考文献

- [1] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick, "Mask R-CNN," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp.2961-2969, (2017).
- [2] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, C. Lawrence Zitnick, "Microsoft COCO: Common Objects in Context," ECCV 2014, pp.740-755, (2014).
- [3] Alexander Mathis, Pranav Mamidanna, Kevin, M., Cury, Taiga Ave, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, Matthias Bethge, "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," Nature Neuroscience 21, pp.1281-1289, (2018).

本プロジェクトに関する業績

- 1) 上原祐介, 大谷直也, 西本圭佑, 川崎邦将, 松井康浩, 長田茂美, 北村俊平, 矢田豊, "深層学習に基づく野生動物認識システム," 中部森林研究, 第68号, pp.25-30, (2020).