

ウミネコ動画の自動分析

井出 佑汰^{1,a)} 水谷友一^{2,b)} 依田 憲^{2,c)} 柳井 啓司^{1,d)}

概要

近年、計算機の能力向上により深層学習による動画中の物体の検出が容易になってきて自動運転や農業や工場などにもその技術は使われ始めている。現在、生物学者は動物の生態観察を長時間にわたる動画を人の目で見ることによって記録付けを行っていて、その作業はかなりの時間を要する。そこで本研究では、最新の深層学習に基づく物体検出手法 [6] と物体追跡の手法 [7] を用いてウミネコの生態観察の自動化を試みた。

1. はじめに

近年、計算機の能力向上により深層学習による動画中の物体の検出が容易になり、世の中に広まりつつある。最近では、自動車の自動運転に深層学習による物体検出の技術が用いられています。注目を受けている。それだけではなく、農業や工場での品質管理にも利用されていて、使用される業界も多方面にわたっている。深層学習を用いた物体検出の代表例としては、SSD[4] や YOLOv3[6] などが有り様々な手法が日々開発されている。本研究で最新の物体検出の手法である YOLOv3[6] を用いて動画中のウミネコの検出を行う。YOLOv3[6] は高速かつ高精度での物体検出が可能となっている。深層学習による手法として物体追跡もある。一連の動画の中で特定の物体を追跡するもので、近年では監視カメラや画像認識と合わせて車載カメラにも用いられている。物体追跡の手法の代表例として High Performance Visual Tracking with Siamese Region Proposal Network[3] や Fast Online Object Tracking and Segmentation: A Unifying Approach[7] がある。本研究では、YOLOv3 で検出したウミネコを追跡する目的として用いている。物体検出の手法と物体追跡の手法を合わせる。

現在、生物学者は動物の生態観察を長時間にわたる動画を実際に人の目で見ることによって記録付けを行っている。その作業はかなりの時間と労力を要する。

そこで本研究では、最新の CNN に基づく物体検出の手法 [6] と物体追跡の手法 [7] を組み合わせて用いてウミネコの生態観察の自動化を試みた。今回の研究では、ウミネコが巣の周りにいるのかいないのかに限定して実験を行った。

2. 関連研究

2.1 物体検出

CNN に基づく物体検出の研究として YOLOv3[6] がある。これまでの物体検出手法である YOLOv2[5] では 19 層のモデルであったのにたしてこの論文の手法では層を 53 層にした。これまでは分類問題に Softmax 関数が使われていたが代わりにロジスティック回帰により分類を行った。これらの変更により速度は少し低下したものの既存の物体検出手法よりも高い精度での検出が可能になった。

2.2 物体追跡

CNN に基づく物体追跡の手法として Fast Online Object Tracking and Segmentation: A Unifying Approach[7] がある。この論文では、既存の CNN に基づく物体追跡の手法の Siamese-FC の、物体追跡の性能に加えて、セグメンテーションマスクを生成することを可能にした手法の提案をしている。リアルタイム性を維持したまま VOT2018 で高い精度を出している。ネットワーク図を図 1 に示す。

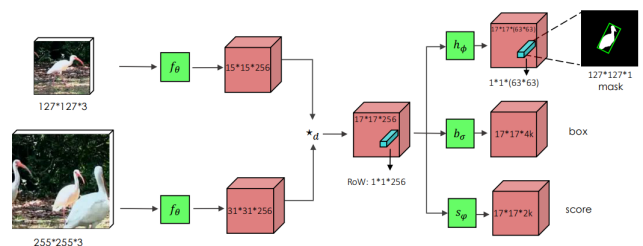


図 1 Fast Online Object Tracking and Segmentation の概要 ([7] より引用)

2.3 動物動画解析

動物動画解析の研究として Recognition in Terra Incognita[1] がある。この論文では、近年の CNN に基づく手法は、学習したデータセットと同じ場所のデータに対して活用することで優れた性能が示されるため解析する対象に対

¹ 電気通信大学 情報理工学研究所 情報学専攻

² 名古屋大学 環境学研究所 都市環境学専攻

a) ide-y@mm.inf.uec.ac.jp

b) yuichi-san@nagoya-u.jp

c) yoda.ken@nagoya-u.jp

d) yanai@cs.uec.ac.jp

して適切なデータセットを用いることを述べている。この論文の考え方に従い、本研究で用いた物体検出手法 [6] のファインチューニングを公開済みのデータセットである ImageNet[2] の black-tailed gull を用いずに、水谷らから提供された燕島のウミネコデータセットにて行った。データセットの詳細については 3 データセットにて説明する。

3. データセット

3.1 評価用データセット

本研究で解析するデータは水谷らから提供された定点カメラにて観測された、ウミネコの燕島での 1 日 (2015 年 5 月 28 日、4:00 ~ 18:00) の動画である。図 2 に示すようにウミネコの巣が A ~F まであり、その周囲をウミネコは行き来する。本研究では、各巣においてどの時間帯にウミネコが存在したのかを記録することを目的とする。



図 2 水谷らから提供されたデータ (2015 年 5 月 28 日分)

3.2 学習データセット

学習用データセットとして、燕島での別の日の 1 日 (2018 年 5 月 29 日、6 月 1 日) のものを使用した。また以下の図 3 に示すように評価用のデータセットとは異なるアングルから撮影されたものとなっている。



図 3 水谷らから提供されたデータ (2018 年 5 月 29 日分)

4. 提案手法

動画を解析するにあたり図 2 の状態では各巣 A ~F の周りにいるウミネコを詳細に記録することができないため、各巣 A ~F の周りで動画を切り取り解像度を下げてから本研究を行った。

本研究は以下のような流れになる。

- (1) 物体検出手法によるウミネコの検出
- (2) 検出された記録をもとにして物体追跡

5. 手法の詳細

5.1 物体検出

本研究の物体検出は YOLOv3[6] を水谷らから提供された学習用データセットでファインチューニングしたもので行う。学習用データセットを評価するときの巣の大きさと同じ解像度の大きさに元の動画を切り取った。さらに、切り取った動画を画像にしてファインチューニングを行った。学習に使用した画像は 4562 枚で検出するクラスは本研究での対象であるウミネコクラスのみとした。

5.2 物体追跡

物体追跡は SiamseMask[7] を用いた。YOLOv3[6] で得られたバウンディングボックスの座標を入力として追跡を行った。物体検出に物体追跡を加えることにより物体検出で検出できなかったフレームを補うことができ、精度の向上が見られると考え本手法を用いた。

6. 実験の結果

6.1 評価方法

本研究の評価は水谷らから提供された実際のウミネコ観察記録を用いて行う。観察記録には評価用のデータセットに対応した観察記録が載っている。評価用データセットに対して本研究の手法を用いて精度を測る。以下図 4 に実際の観察記録の一例を載せる。

時間	座標(北緯/東経)	姿勢(立/座)	状態	行動(飛/歩)	注1	注2
4:00:04		立	無	無		
(8) 5:06:16		立	無	無	0 (おそらく) つがいでない個体が接近	(8) 巣の近くに長く居たので記録しませ
(8) 5:08:09		立	無	無	0 (おそらく) つがいでない個体が接近	(8) 巣の近くに長く居たので記録しませ
5:27:05		立	有り	0	0 転倒 (ウ)	
6:52:33		立	有り	0	0 転倒 (ウ)	
7:36:08		立	有り	0		
8:15:21		観察者の侵入				
11:19:18		立	有り	0		
12:58:12		立ち上がる				
14:30:39		座席に寝かす				
14:56:42		0			1 が離脱	
14:59:54		立	無	0	1 が帰巢	
14:30:07	142	立	無	0	2 が帰巢	
14:30:25	2	座	無	0	1 が離脱	
14:56:19	2	立	有り	0		
15:01:59	2	立	有り	0		
15:49:16	2	立	有り	0		
16:17:19	2	立	有り	0		
17:22:58	2	立	有り	0		
17:59:31	2	立	有り	0		

図 4 実際の観察記録

6.2 実験の結果

各巣の結果を以下の表 1 に示す

表 1 本手法を用いた正解率 [%]

巣	A	B	C	D	E	F
検出のみ	86.4	45.0	19.0	60.2	52.8	20.6
検出+追跡	88.6	80.0	29.3	64.1	50.9	70.6

7. 考察

7.1 巣 A について

ウミネコの体全体がよく映っているため検出そのものがうまくいっているフレームが多く見受けられた。下の図 5 のようにウミネコクラスである確率そのものも高いものになっている。そのため、誤検出も少なかった。検出そのものが上手くいっているため追跡の機能を追加したときに精度が向上した。



図 5 巣 A の様子

7.2 巣 B について

巣 B 以下の図 6 のようになっていて他の巣と比べると開けている。そのためウミネコの全体像が見えやすくなっているため検出そのものはうまくいくこともあった。そのため、追跡もうまく機能して精度が向上したと考えられる。しかしながら巣の手前を他の巣のウミネコが動き回るため、奥にいるウミネコの姿が隠れてしまい検出されないこともあった。追跡をする際に巣の手前を歩いている別の巣のウミネコが対象となってしまうこともあり、動画の切り取り方の工夫が必要であると考えられる。



図 6 巣 B の様子

7.3 巣 C について

巣全体が草の中に埋もれているため、ウミネコがいたとしても物体検出器で検出そのものが非常に困難であった。以下の図 7 の場合、実際には真ん中の白く見えるところにウミネコが存在しているのだが人の目で見ても見落としてしまうくらいには見分けるのが困難であるといえる。巣 B と同様に手前の別の巣のウミネコを検出して追跡してしまうことがあったため、巣の切り取り方を変える必要があると考えられる。



図 7 巣 C の様子

7.4 巣 D について

巣 D では以下の図 8 のような誤検出が見られた。学習データの中に雛の画像は用いていないのだが、今回の実験では、雛もウミネコと認識された。雛を検出するとその後雛を追跡してしまう。そのため、ウミネコの全体像が比較的良好に見えていたのにも関わらず精度が高くならなかったと考えられる。



図 8 巣 D での誤検出

7.5 巣 E について

草が生い茂っている中に巣があるためウミネコが巣で座っていたとしても検出されないケースや以下の図 9 のように巣の手前の岩をウミネコと誤検出してしまうフレームが多く見受けられた。誤検出があるとその後の追跡において、本来ウミネコを追跡することを目的としているにもかかわらず他の物体を追跡してしまう。そのため検出のみの場合よりも検出と追跡を合わせた手法のほうが低い精度となったと考えられる



図 9 巣 E の誤検出の様子

7.6 巣 F について

以下の図 10 のように巣 F は草が生い茂っているところに巣があるため、ウミネコそのものの検出が難しくなっている。一方で巣 C よりもウミネコの体全体が見えるので巣 C よりも高い精度となった。検出が出来ているフレームがあることから追跡の機能が上手く働いた。そのため検出と追跡を合わせたところ精度が大幅に向上したと考えられる。



図 10 巣 F の様子

7.7 物体検出と物体追跡を組み合わせた効果について

以下の図 11 はフレームとフレームの間における物体検出器の未検出フレームを物体追跡の手法を組み合わせたことにより補う様子が示されている。青い枠が物体検出によるもので、緑色の枠が物体追跡によるものとなっている。最初と最後のフレームでは正しく検出できているにもかかわらず、途中のフレームにおいては検出できないというのは、物体検出のみを使用していると起こることである。物体追跡の手法を組み合わせることで未検出フレームを減らすことが可能になり、物体検出手法のみの場合よりも正確な観察記録を付けることができるといえる。

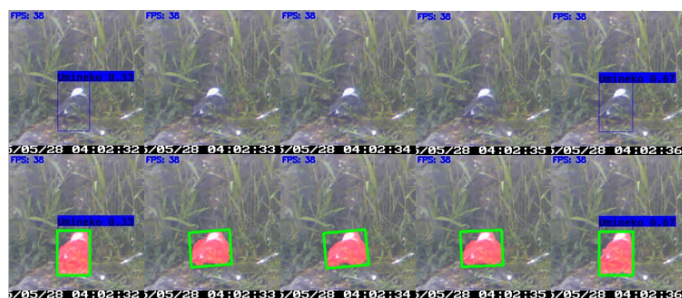


図 11 未検出のフレームを補う様子上：検出のみ 下：検出+追跡

8. 今後の課題

本研究では、ウミネコがいるのかいないのかについてのみしか記録することができなかった。全体像が十分に見ることのできない場合における検出率が低く、検出できないことそのものが精度に大きくかかわってくるため、学習データを増やすことによりさらなる精度向上が必要であるといえる。巣の切り取り方が検出にどの程度影響を与えるのかについて検証する必要がある。物体検出と物体追跡の手法の組み合わせにより未検出のフレームを減らすことが出来た。そのことが今回の観察記録の精度向上につながった。しかし、ウミネコ巣から出て行ってフレームアウトした際の対応について考える必要がある。実際の観察記録では姿勢や転卵なども記録されている。完全な自動分析にするには検出だけでなく、検出と同時に行動も推定するシステムが必要といえる。検出したフレームに対しての動作推定の機能を追加していく必要がある。

参考文献

- [1] Beery, S., Horn, G. V. and Perona, A. P.: Recognition in Terra Incognita, *Proc. of European Conference on Computer Vision* (2018).
- [2] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L.: Imagenet: A large-scale hierarchical image database, *Proc. of IEEE Computer Vision and Pattern Recognition*, Ieee (2009).
- [3] Li, B., Yan, J., Wu, W. and Zhu, Z. and Hu, X.: High Performance Visual Tracking with Siamese Region Proposal Network, *Proc. of IEEE Computer Vision and Pattern Recognition* (2018).
- [4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C.: Ssd: Single shot multi-box detector, *European conference on computer vision*, Springer, pp. 21-37 (2016).
- [5] Redmon, J. and Farhadi, A.: YOLO9000: better, faster, stronger, *Proc. of IEEE Computer Vision and Pattern Recognition* (2017).
- [6] Redmon, J. and Farhadi, A.: Yolov3: An incremental improvement, *arXiv preprint arXiv:1804.02767* (2018).
- [7] Wang, Q., Zhang, L., Bertinetto, L., Hu, W. and Torr, P. H.: Fast Online Object Tracking and Segmentation: A Unifying Approach, *Proc. of IEEE Computer Vision and Pattern Recognition* (2019).