

## 科学研究費助成事業 研究成果報告書

平成 30 年 6 月 5 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2016～2017

課題番号：16K16083

研究課題名(和文) 深層学習を用いた動画からの物体の検出と分類

研究課題名(英文) Moving object detection and classification using deep learning

研究代表者

川上 玲 (Kawakami, Rei)

東京大学・大学院情報理工学系研究科・特任講師

研究者番号：90591305

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本事業では、動画において、動きでしか種類が分からないほど小さく写る物体の検出と分類に取り組んだ。野鳥の広域監視を出口として、風車周辺の海ワシ類を対象に4Kの解像度の動画収集を行い768GBの動画に対し鳥の軌跡や背景の部分のラベル付が完了した。検出・分類の手法として、最終的に深層特徴の相関フィルタを用いた追跡と畳込LSTMによる動きの学習により、物体を追跡しながら動きパターンにより鳥か否かを識別する手法の開発につながり、これが静止画のみに基づくベースラインから25.2%ポイントの性能向上を得た。これらの成果は多くの論文として発表されたほか、風車による野鳥の生態系への影響の軽減につながる。

研究成果の概要(英文)：In this project, we worked on the detection and classification of objects that are too small that they can only be recognized by their motion. With wide area surveillance of wild birds as a target application, we collected 4K resolution videos for sea eagles around the wind turbines and labeled bird trajectories and background parts for 768 GB videos. As a method of detection and classification, tracking using a correlation filter of deep feature and motion learning by convolutional LSTM leads to a method that identifies a bird by its motion patterns acquired by simultaneously tracking them. It achieved a performance improvement of 25.2% points from baselines that only use still-image features. The achievements appeared as several papers, and they will contribute to reduce the impact of wind turbines on the ecology of wild birds.

研究分野：コンピュータビジョン

キーワード：動き 学習 深層 鳥 風力発電 物体検出 分類

### 1. 研究開始当初の背景

深層学習 (Deep Learning) がコンピュータビジョンや音声認識の分野で従来の性能を高く上回る成果を上げ注目を集めている。画像分類のタスクにおいては、深層学習に基づく手法の誤差率が性能が良いものの、検出のタスクは、引き続き改善の余地がある。(分類は限られたクラス内から正解を導くが、検出は負例が限りなく存在するため、正解を導くことがより難しい。)

2015年から、動画を用いた認識が再注目されている。ILSVRC2015では動画による物体検出用の学習データが公開された。動き情報は、古くから物体の検出、追跡や、認識に用いられてきた。しかし近年は、毎フレーム検出器を走査させる手法が検出性能が良く一般的で、動きを認識に用いる研究事例は(行動認識を除けば)少ない。行動認識では動きを用いた研究が盛んであるが、解析が容易な動画セットは深層学習用にはサンプルが少なく、比較的大規模なものは、ゴルフをする、フェンシングをする、など、動画が複雑で、動き情報と静止画の情報を分離した解析が難しく、動き特徴の明確な理解に至っていない。深層学習は数万という桁の学習データを要するため、これを作成するコストが研究の障害となる。ILSVRC2015で新たに動画による物体検出用学習データが公開され、今後数年でさらなるデータの増加が期待できることから、動画による認識を再考する良い機会である。

### 2. 研究の目的

申請者はこれまで、風力発電所の周囲の鳥画像認識の課題に取り組んできた。風力発電所ではブレードにワシなどの希少種が衝突死するバードストライクが深刻化しており、鳥の自動認識が急務である。人間がラベルデータを作成する際は、解像度が低く鳥の判別が難しい場合でも、飛び方により識別を行うことが可能である(図1)。動き情報の活用により鳥の検出や識別の性能の向上が確実に見込まれ、深層学習による動きの学習は学術的にも興味深いことから、静止画特徴では判断できない低解像度の動物体の検出や識別の課題に取り組む。



図1：飛び方により種の判別が可能な例

### 3. 研究の方法

本事業を通じ、三つの動きを使った認識手法を開発した。

(1) 初年度に取り組んだのは、フレーム間差分の情報から検出率を向上させる手法で

ある(一つ目)。二フレームの情報でも動物体と静止物体は充分区別できることが多く、フレーム間差分は高速に計算できるため、この情報を用いて深層学習を行い、歩行者検出のデータセットで性能を確認した。動画における「動き」は、カメラの動き、対象物の移動、対象物の部分の動き、に分類できるが、検出に役立つ動き特徴は部分の動きである。そこで、部分の動きを粗い optical flow で抽出し、部分の動きの差分画像と静止画情報の二つをそれぞれ Convolutional Neural Networks (CNN) に学習させ、二つの CNN を Boosted Forest で統合し、背景と対象物の分類を行わせた。差分画像の学習には行動認識で用いられる動画データベースを用い、これが有効であることを確認した。静止画特徴のみを用いる手法と比較して、歩行者検出のベンチマーク (Caltech Pedestrian) で 2.1% ポイントの性能向上を確認した。

(2) フレーム間差分の手法と並行して、最適な動き特徴に必要なフレーム数を決めずに、Recurrent Neural Networks (RNN) の一つである Long Short-term Memory networks (LSTM) を用いて、特徴やフレーム数を自動的に学習する手法の開発に取り組んだ。初年度では、最大5フレームの入力から全結合型 LSTM を用いて動きを学習し、対象物とそれ以外の分類を行う手法を開発した(二つ目)。静止画特徴量のみを用いる手法と比較し、4%ポイントの性能向上を確認した。

(3) 動物体の認識を行う際、(行動認識などの動画の分類と違って)画像のどこからどこへ対象物が移動したかが分からなければ、その物体の動きを機械に学習させることはできない。したがって、動物体の認識を行うには、追跡も同時に解く必要がある。初年度は、既存の追跡器を改良して前処理として用いたが、二年目は追跡と検出を同時に行う手法を開発した(三つ目)。前年度の LSTM の部分を画像の局所特徴の学習により適した 畳込 LSTM に拡張し、追跡は特徴量学習で得た特徴マップ上での相関フィルタにより行った。結果として、物体を追跡しながらその動きパターンにより対象物体か否かを識別させることに成功した。静止画特徴量のみを用いる手法と比較し、同じ深さの深層学習を使った場合でも 25.2%ポイントの性能向上を得た。

### 4. 研究成果

(1) ~ (3) 深層学習用データとして、共著者や風車を管理する公共機関との協力の下、北海道苫前町において約一ヶ月に渡り風車周辺の花ワシ類を対象として 4K の解像度の動画収集を行い、このうち三日分のデータ 768GB に対して飛行する鳥の軌跡や背景の部分に対するラベル付が完了した。図2に動画記録装置を示す。データは出版された論文

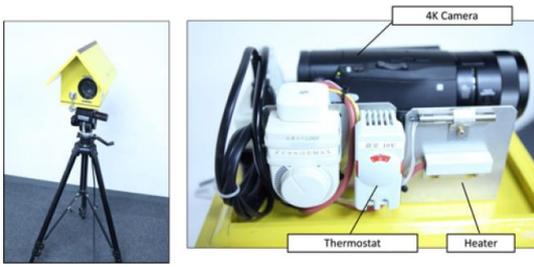


図 2 : 動画記録装置

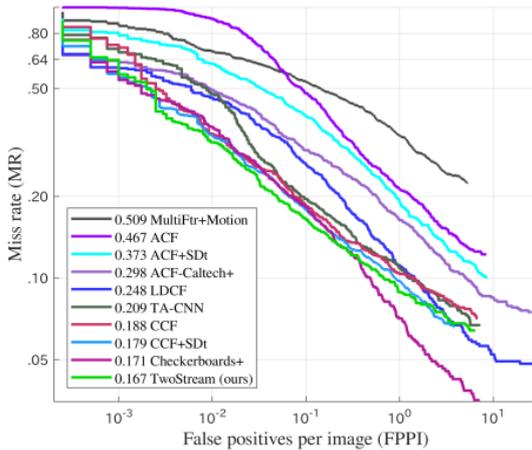


図 3 : フレーム間差分による検出手法の性能



図 4 : フレーム間差分の手法が成功している例. 赤いボックスが深層特徴と Boosted Forest の検出結果, 青が提案手法の検出結果, 黄色が真の人領域である.

で使用された.

(1) フレーム間差分の手法では, 人検出のベンチマーク (Caltech Pedestrian) において, 開発当時 (2016 年) で世界三位の認識性能を達成した (average MR=16.7%). 図 3 に性能の詳細を示す. 縦軸が見落とし率 (Miss Rate), 横軸が画像一枚当たりの誤検出数 (False positives per image) である. 横軸を変化させたときの縦軸の平均が average MR であり, 凡例の数値がこれにあたる. CCF (深層特徴量と Boosted Forest の組み合わせ) と比較して, 提案手法は 2.1%ポイントの性能向上が見られる. また単純な動きの学習 (CCF+SDt) では average MR が 17.9%であったので, 提案手

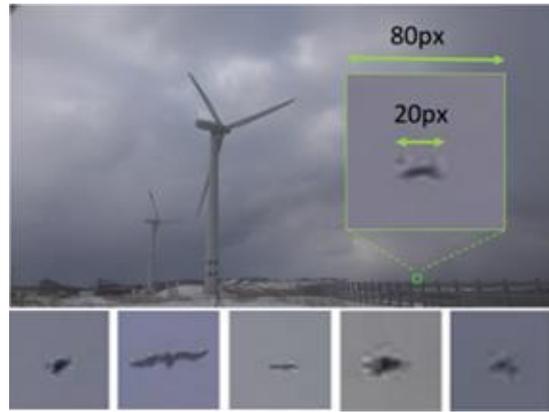


図 5 : 申請者らの動画データベースの鳥

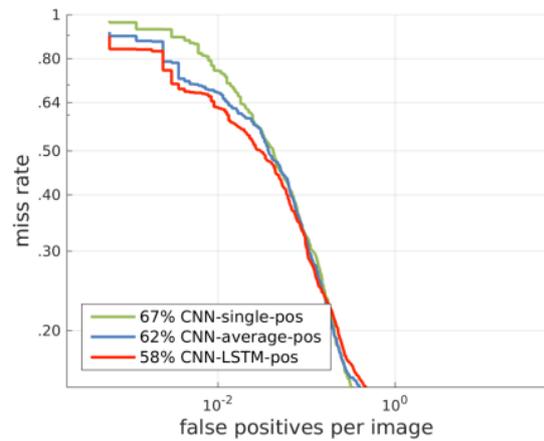
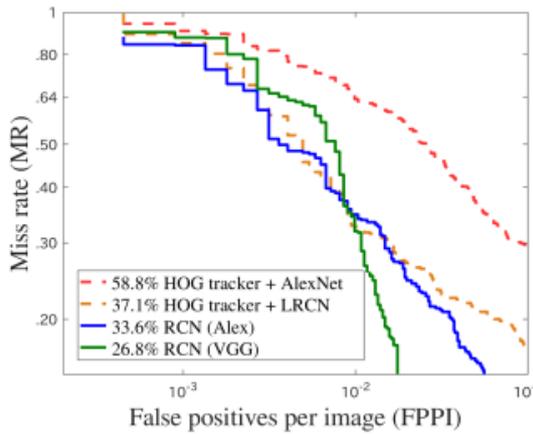


図 6 : 全結合型 LSTM による鳥の検出性能

法はこれよりも 1.2%ポイント向上できることが分かる. 図 4 に提案手法の成功例を示す. この成果は英文論文誌に投稿中であり, 2018 年 5 月に Major revision を投稿した.

(2) 全結合型 LSTM の手法では, 申請者らが作成した鳥の動画のベンチマークで評価した. ベンチマークの画像例を図 5 に示す. 静止画のみに基づく手法 (全フレームに渡って CNN の出力を平均する) と比較し, 2016 年時点で数%ポイントの性能向上を達成した. 性能は解像度によって異なるが, すべての解像度を含んだ実験では, 図 6 に示すように, CNN の平均 (CNN-average) が average MR で 62% であったのに対し, 提案手法 (CNN-LSTM) では 58% であり, 4%ポイントの向上が見られる. この成果は風力発電の国際会議に採択された.

(3) 追跡と検出を同時に行う手法では, 申請者らの鳥動画のベンチマークと, 既存のドローンの飛行動画で評価した. 鳥の検出性能を図 7 に示す. 図 7 は, 図 6 に対応するものであり, HOG tracker + AlexNet が既存の追跡器と静止画特徴のみに基づく手法 (図 6 における CNN-average), HOG tracker + LRCN が (2) の改良 (図 6 における CNN-LSTM) に当たる. 静止画特徴のみに基づく手法, (2)



Reasonable subset (40 pixels -)  
 図7：検出と追跡の同時推定による鳥の検出性能

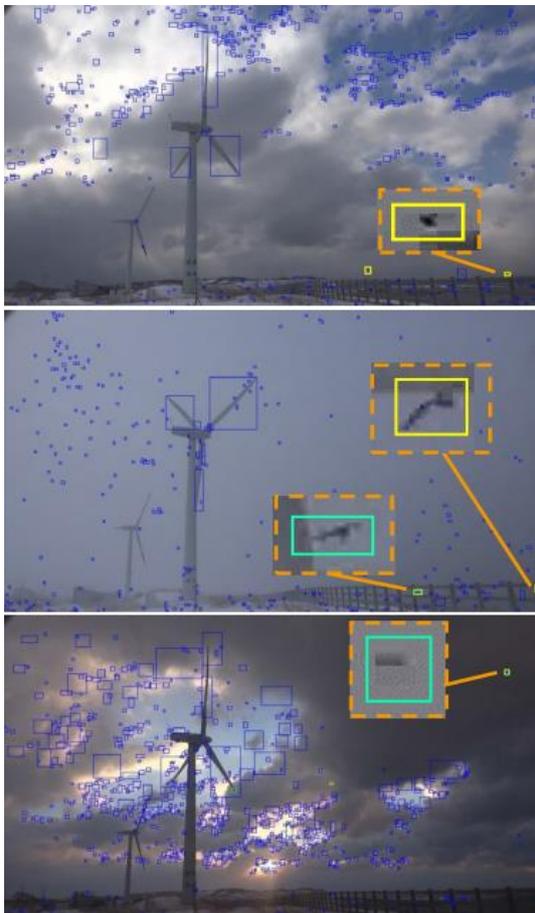


図8：検出と追跡の同時推定による鳥検出の例。オレンジ色の囲みは真の鳥領域を表し、中のボックスが検出器で鳥のスコアが高いものである。黄色～緑色のボックスは提案した検出器でスコアが高いもの、青色は低いものを表し、背景差分により検出された候補領域には全てスコア付けがなされている。画像サイズは4Kであり、鳥領域は20ピクセル程度である。

の改良、の双方とも図6と比較して average MR が向上しており、これは主に事前学習の利用とパラメータ調整による。さらに提案し

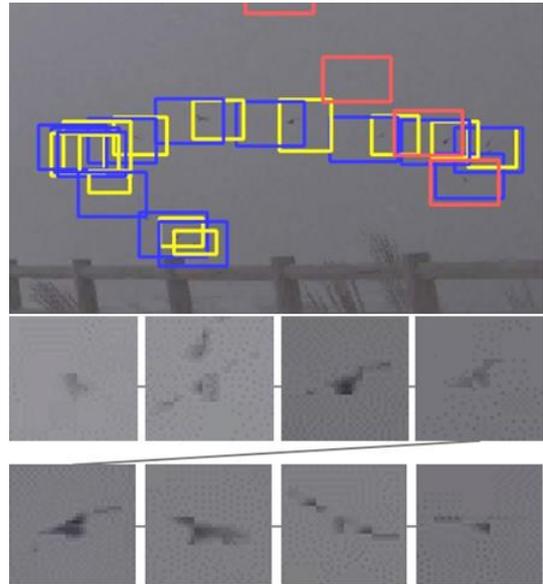


図9：鳥追跡の例。上の画像において、青が提案した追跡器による結果、黄色が真の軌跡、赤が HOG 特徴量による追跡器の結果である。下の画像列は、提案手法により追跡した画像の列である。鳥検出で学習した特徴量を用いているために、提案手法は低解像度の鳥領域を頑健に追跡できる。

た手法である追跡と検出の同時推定（図6のRCN）により、average MRは26.8%にまで向上した。静止画特徴のみに基づく手法（HOG tracker + AlexNet）から比較して同じ深さのCNNを使ったRCN（Alex）でも25.2%ポイントの向上を得た。

検出した画像例を図8に、追跡結果の例を図9に、それぞれ示す。図8に示すように、静止画のみでは、鳥かどうか人間にも判断が難しい画像でも複数フレームに渡る動きの利用により鳥の検出に成功している。また、図9から、既存の追跡器ではすぐに見失ってしまうような鳥の追跡に成功していることが分かる。この成果は、arxivに投稿済みであり（[その他]のリンク参照）、英文論文誌への投稿を準備している。

## 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 2 件）

- ① R. Yoshihashi, R. Kawakami, M. Iida, T. Naemura, “Bird detection and species classification with time-lapse images around a wind farm: Dataset construction and evaluation.” Wind Energy, Wiley, 査読有, 20(12): 1983-1995, 2017, DOI: 10.1002/we.2135.
- ② A. Takeki, T. T. Trinh, R. Yoshihashi, R. Kawakami, M. Iida, T. Naemura, “Combining deep features for object

detection at various scales: Finding small birds in landscape images." IPSJ Transactions on Computer Vision and Applications, 8(5), 査読有, 2016, DOI 10.1186/s41074-016-0006-z.

[学会発表] (計 8 件)

- ① Tu Tuan Trinh, Ryota Yoshihashi, Rei Kawakami, Makoto Iida, Takeshi Naemura, "Bird Detection near Wind Turbines from High-resolution Video using LSTM Networks." World Wind Energy Conference and Exhibition (WVEC2016), Tokyo, October 2016.
- ② Akito Takeki, Tu Tuan Trinh, Ryota Yoshihashi, Rei Kawakami, Makoto Iida, Takeshi Naemura, "Detection of small birds in large images by combining a deep detector with semantic segmentation." In Proc. of IEEE International Conference on Image Processing (ICIP2016), 2016.
- ③ 邵文, 福田誠一郎, 吉橋亮太, 川上玲, 尤少迪, 川瀬英路, 苗村健, "放牧支援のための空撮画像における CNN に基づく牛検出." 映像メディア処理シンポジウム (IMPS2017) P-1-6 (2017.11).
- ④ 王晋, 福田誠一郎, 吉橋亮太, 川上玲, 川瀬英路, 苗村健, "畜産業支援に向けたドローンによる空撮画像の撮影と牛検出への応用." 第 23 回画像センシングシンポジウム SSII, IS1-07, June 2017.
- ⑤ チン トゥアントゥー, 吉橋亮太, 川上玲, 尤少迪, 飯田誠, 苗村健: "動物体検出のための LSTM ネットワークによる静止画と動き情報の統合", 信学技報 PRMU2016-233, vol. 116, no. 528, pp. 221 -- 226 (2017.3).
- ⑥ 福田誠一郎, 吉橋亮太, 川上玲, 飯田誠, 苗村健: "鳥検出のための矩形領域を教師情報とした CNN による領域分割", 映像情報メディア年次大会, 24D-4 (2016.09).
- ⑦ 竹木章人, チン トゥアン トゥー, 吉橋亮太, 川上玲, 飯田誠, 苗村健: "深層特徴量の組み合わせによる解像度に適応的な物体検出 —遠景画像に小さく写る鳥の検出—", 画像の理解・認識シンポジウム (MIRU2016), PS2-05 (2016.8)
- ⑧ Tran Hoang Ba, 吉橋亮太, 川上玲, 飯田誠, 苗村健: "複雑な背景や変形を考慮したロバストな鳥の自動追跡", SSII16, IS2-35 (2016.6).

[図書] (計 件)

[産業財産権]

○出願状況 (計 件)

名称 :  
発明者 :  
権利者 :  
種類 :  
番号 :  
出願年月日 :  
国内外の別 :

○取得状況 (計 件)

名称 :  
発明者 :  
権利者 :  
種類 :  
番号 :  
取得年月日 :  
国内外の別 :

[その他]  
ホームページ等  
研究代表者HP  
<http://www.nae-lab.org/~rei/research/research.html>

検出と追跡の同時推定  
<https://arxiv.org/abs/1709.04666>

6. 研究組織  
(1) 研究代表者  
川上 玲 (KAWAKAMI, Rei)  
東京大学・大学院情報理工学系研究科・特任講師  
研究者番号 : 90591305

(2) 研究分担者 ( )

研究者番号 :

(3) 連携研究者 ( )

研究者番号 :

(4) 研究協力者 ( )