H-023

Attention の導入による YOLOv5 と DeepSORT の改良に基づく放牧牛のリアルタイム追跡手法の構築

A real-time tracking method of grazing cattle based on improved Deep SORT and YOLOv5 based on attention mechanism

楊陽† 小松 瑞果† 大山 憲二† 大川 剛直†
Yang Yang Mizuka Komatsu Kenji Oyama Takenao Ohkawa

1. はじめに

近年，畜産分野において，農業従事者の減少や高齢化が大きな問題となり，情報技術の導入による，作業の効率化や自動化の必要性が叫ばれている．我々の研究グループでは，農業従事者の負担軽減を目的として，カメラによる放牧牛の自動モニタリング手法について開発を進めている．その実現のため，カメラ画像上で牛を特定し，その行動をリアルタイムに追跡する必要がある．多物体追跡(Multiple Object Tracking: MOT) は，動画とオブジェクトのクラスが与えられたときに，動画中のオブジェクトの位置を特定しつつ，個別のオブジェクトがどのように移動したかを追跡するタスクである．近年，深層学習技術の進捗により，MOT アルゴリズムが多数提案されている[4].

MOT アルゴリズムに採用されている標準的なアプローチは，Tracking-by-Detection である． Tracking-by-Detection では，まず入力画像に対して，物体検出を行い，検出された物体を動画フレーム間で関連づける．

本研究では，俯瞰動画像に基づく放牧牛のリアルタイム追跡に焦点を当て，そのための手法を提案する．提案手法は，YOLOv5[5]を検出器として利用し，MOT の一つである DeepSORT[1] に基づく． Attention を導入することで YOLOv5 を改良し，検出精度を向上する．また，追跡においては，改良される YOLOv5 にて抽出された特徴を再利用することで，処理速度を向上させる．神戸大学大学院農学研究科附属食資源教育研究センターにおいて取得された動画像に対して提案手法を適用し，その有効性を示す.

2. Tracking-by-Detection による物体追跡

2.1 物体検出

物体検出とはコンピューター・ビジョンにおける一つのタスクであり，画像内の特定の物体位置を特定する手法である．物体検出のアルゴリズムは，物体を直接検出する 1 段階法と，まずバウンディングボックスの候補を複数生成し，それぞれ物体認識を行う 2 段階法に大別される．YOLO シリーズなどは 1 段階法であり，実行速度が速い．一方で，RCNN[3]などの 2 段階法は，1 段階法と比べて検出精度は高いが，実行速度が遅い．本研究が扱う問題では，物体検出をリアルタイムに実行することが望まれ，YOLOv5 を検出モデルとして利用した.

2.2 物体追跡

本研究では，関連度の評価(図 1， Stage 4)において，物体の運動情報と外観特徴を用いる DeepSORT により物体追を行う(図 1 参照).

図 1　DeepSORT に基づく物体追跡の手順

具体的には，カルマンフィルターを用いて予測された物体の軌跡と，事前に検出された物体のバウンディングボックスの関連度を算出する．また，CNN を用いて抽出された外観特徴（図 1，Stage 3）と，バウンディングボックスの関連度を算出する．

関連付け(図 1，Stage 4)においては，ハンガリアンアルゴリズムを用いる．具体的には，算出された二つの関連度の重み付き和を求め，これをコスト関数とみなす．これにより，検出されたバウンディングボックス追跡番号に関連づける.

2.3 Attention 機構

近年，深層学習技術の一つとして Attention 機構が注目されている．Attention 機構は様々なニューラルネットワークに取り入れることができ，様々なタスクにおいて，その重要性が確認されている．

CBAM(Convolutional Block Attention Module)[2] は，Attention メカニズムの一手法であり，Attention 構造を組み込んだ畳み込みニューラルネットワークの導入により，より有用な特徴マップを得られる．CBAM は軽量であるため，様々な CNN に組み込むことができ，End-to-End な学習が可能である．本研究では，物体検出と特徴抽出（図 1， Stage 2， 3）において，これを導入する.

3. 提案手法

Tracking-by-Detection の物体追跡の精度は，物体検出の精度に強く依存する．本研究における追跡対象である牛に関して，色彩特徴と画像中における大きさから，オリジナルの YOLOv5 では十分な追跡精度が得られない．そこで，本研究では，牛の特徴の強化のために CBAM を導入し，YOLOv5 モデルと組み合わせる．一般に，畳み込みニューラルネットワークについて，深い層で抽出される特徴ほど抽象的かつ複雑であり，表現能力が高いと言われている．これを活用するため，本研究では，YOLOv5 の最後の出力層の直前にのみ，CBAM ブロックを導入する（図 2 参照）

† 神戸大学 Kobe University

このような導入方法により，計算コストを抑えつつ，特徴量が強化されると期待できる．

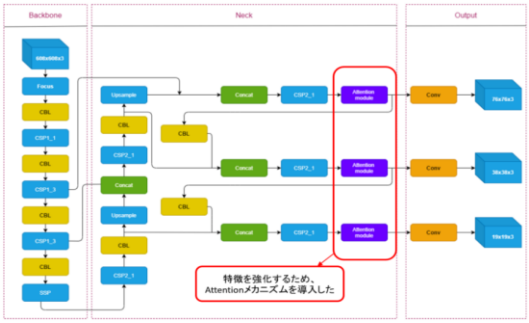

図２　改良した YOLOv5

このような構成により，CBAM による計算コストはある程度抑えられるものの，追跡のリアルタイム性を保証するためには，さらなる工夫が必要である．そこで，本研究では空間方向とチャンネル方向両方の attention の軽量化を行う．従来の CBAM では，まず，チャンネル方向への，CBAM は入力特徴マップ$F \in R^{C\times H\times W}$ に対し，最大プーリングと平均プーリングを適用することで，二つの特徴を得る．さらに MLP を用いて，式(1)のようにチャンネル attention マップ$M_c \in R^{C\times 1\times 1}$ を生成する．

$$M_c(F) = \sigma\left(MLP\left(Avgpool(F)\right) + MLP\left(Maxpool(F)\right)\right) = \sigma\left(W_1\left(W_0\left(F^c_{Avg}\right)\right) + W_1\left(W_0\left(F^c_{Max}\right)\right)\right) \quad (1)$$

ここで，σ はシグモイド関数を表し，W_1 と W_0 は MLP の重みを表す．この場合，MLP の全結合層のパラメータ数は，入力特徴マップのチャンネル数の二乗に比例することになる．この点を踏まえ，本研究では，MLP を次式のようにカーネルサイズが 1×K の 1 次元畳み込み層に代替した．

$$M_c(F) = \sigma\left(f^k_{1D}\left(Avgpool(F)\right) + f^k_{1D}\left(Maxpool(F)\right)\right) = \sigma\left(f^k_{1D}\left(F^c_{Avg}\right) + f^k_{1D}\left(F^c_{Max}\right)\right) \quad (2)$$

ここで，K はチャンネル数を表す．また，チャンネル方向に強化された特徴 $F' = M_c(F)\,F$に対して，空間方向の特徴強化を行う．ここで，従来の CBAM では，式(3)のように最大プーリングと平均プーリングを用いた後に，カーネルサイズを 7 × 7とする畳み込み層を適用し，空間 attention マップ$M_s(F) \in R^{H\times W}$ を生成する．

$$M_s(F) = \sigma\left(f^{7\times 7}\left([AvgPool(F'); Maxpool(F')]\right)\right) = \sigma\left(f^{7\times 7}\left(F^s_{Avg}; F^s_{max}\right)\right) \quad (3)$$

7 × 7 カーネルは受容野が広いため，計算量も増加する．そこで，本研究では，式(4)に示すように，3 × 3 の dilation 畳み込みカーネルを利用する．

$$M_s(F) = \sigma\left(f^{3\times 3(dilation)}\left(F^s_{Avg}; F^s_{max}\right)\right) \quad (4)$$

さらに，従来の CBAM と同じ大きさの受容野をもたせるべく，本研究では，セル間隔を 2 に設定する．最後に，空間方向に強化された特徴 $F'' = M_c(F')\,F'$を算出する．

追跡において，DeepSORT で検出した物体に対して，再度 CNN を用いることで特徴の再抽出を行う場合，計算コストがかかる．しかし，本研究では CBAM を改良し，特徴の強化を行ったために，特徴量を再度抽出する手順が省略できる．そこで，本研究では，検出モデルで抽出された特徴を再び利用する．

4. 実験と考察

本研究では，神戸大学大学院農学研究科附属食資源教育研究センターにおける黒毛和種繁殖牛(約 30 頭)を対象とし，実験を行う．

物体検出に関する実験として，YOLOv5，YOLOv5 と CBAM，YOLOv5 に attention 機構を導入した提案モデルの，三つのモデルを用いた．ここで，評価指標として mean Average Precision (mAP)を使用した．表 1 に示す実験結果から，提案手法では，精度が YOLOv5+CBAM と同等の精度を維持しつつ，処理速度(fps)が 24 から 30 に向上していることが分かる．

表 1　検出手法間の比較

手法	mAP	FPS
YOLOv5	0.856	38
YOLOv5+CBAM	0.943	24
ours	**0.925**	**30**

追跡に関する実験では，MOT タスクでよく使われる MOTA という指標と処理時間を用いて評価する．結果を表 2 に示す．

表 2　トラッキングの実験結果

手法	MOTA	FPS
YOLOv5+Deepsort	64.2	17
Ours	**71.4**	**25**

表 2 において MOTA が向上していることから，物体検出に導入した attention により対象物体の特徴が強化されたことが示される．また，CBAM の軽量化及び特徴の再利用により，従来手法(YOLOv5+DeepSORT)と比較して，処理速度が向上した．以上により，提案手法により，従来手法より高精度で，かつ，リアルタイム追跡が可能な処理速度である追跡が達成されたことが確認できた．

謝辞

本研究の一部は JSPS 科研費 21H04914 の助成による．

参考文献

[1] N.Wojke, A.Bewley, and D.paulus, “Simple Online and Real-time Tracking with Deep Association Metric.”, *In ICIP 2018*
[2] Woo, Sanghyun, et al. "Cbam: Convolutional block attention Module.", *In ECCV 2018*
[3] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks.", *In NIPS 2015*
[4] Ciaparrone, Gioele, et al. "Deep learning in video multi-object tracking: A survey." *In Neurocomputing 2020*
[5] Jocher, G.: ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements. https://github.com/ultralytics/yolov5