

令和5年度卒業論文

Soft-NMS を追加した SORT による  
重なりに強い複数物体追跡

指導教員 椋木雅之

宮崎大学工学部情報システム工学科

奈須泰樹

## 概要

本研究では、重なりを起こす複数物体追跡において、複数小動物用 SORT を改良することで SORT の追跡の誤り (ID スイッチ) 数の削減を目指す。複数物体追跡手法としてよく用いられる SORT は、物体検出アルゴリズムである YOLO によって検出した追跡対象を、カルマンフィルタで予測した位置とマッチングさせ、追跡 ID を割り振ることで追跡を行う。しかし、SORT では追跡対象の重なりやすれ違いによる検出の失敗から、追跡が途切れてしまい、ID スイッチが起きるという問題がある。そこで、本論文では検出の後処理として行われる NMS を改良した「Soft-NMS」と、重なりに強い複数小動物用 SORT の機能を組み合わせて改良した NasuSORT を提案した。Soft-NMS の導入により検出を行った際、追跡対象の検出漏れを削減することができる。また、複数小動物用 SORT の「二段階マッチング」により、誤った追跡 ID の新規作成を防止できる。

評価実験においては、複数小動物用 SORT で実験を行った小動物の動画に加え、FTC2024 で提供された Sweat Fish データセットを用いて、提案手法と複数小動物用 SORT、SORT の三つで比較した。ID スイッチの数を用いて評価し、結果として、提案手法が ID スイッチの数を削減しつつ、検出数の増加による誤追跡も防止することができた。今後の課題としては、検出が難しい環境での精度向上が挙げられる。検出器をより高性能なモデルに変更し、追跡においても予測位置との IoU だけでなく外観特徴を用いるなど改良することで追跡精度を向上できると考えられる。

# 目次

1	はじめに.....	1
2	複数物体追跡の従来研究.....	3
2.1	複数物体追跡.....	3
2.2	複数小動物用 SORT.....	3
2.3	YOLO.....	6
2.4	NMS.....	6
2.5	Soft-NMS.....	7
3	重なりに強い物体追跡.....	10
3.1	複数物体追跡の問題点.....	10
3.2	Soft-NMS を追加した複数小動物用 SORT.....	10
3.3	提案手法の流れ.....	12
3.3.1	IoU の計算.....	13
3.3.2	Soft-NMS 処理.....	13
3.3.3	二段階マッチング.....	13
4	評価実験.....	14
4.1	実験データ.....	14
4.2	検出器の学習.....	16
4.3	実験設定.....	16
4.4	実験結果.....	17
4.5	実験結果の分析.....	18
4.6	考察.....	22
5	おわりに.....	23
	謝辞.....	24
	参考文献.....	25

## 1 はじめに

複数物体追跡(Multiple Object Tracking: MOT)は、映像の中で移動していく複数の物体を区別して継続的に追跡する技術の総称である。複数物体追跡は、自動運転で周囲の状況を把握するために利用されたり、動物の生態や行動を分析するために自動で記録を取る際に用いられるなど、多種多様な利用方法が期待されている。複数物体追跡手法のひとつである SORT[1]は、深層学習を用いた物体検出器によって検出した追跡対象を、カルマンフィルタで予測した位置とマッチングさせ、追跡 ID を割り振ることで追跡を行う。比較的シンプルなアルゴリズムであり、高速な処理で追跡を行えるため、盛んに利用されるようになってきている。しかし、SORT は検出器から出力された検出結果に対して処理を行うので、追跡対象の重なりや、すれ違いによって一時的に検出ができなくなり、再度検出された際に追跡 ID が別の ID に入れ替わる ID スイッチが発生し、追跡の精度が低下してしまう。従来の複数物体追跡では、人間を対象に重なりが少ない上方からの映像を使って追跡を行うなど検出しやすい状況下で追跡を行っていたので、精度低下は見られなかった。一方で、一般的な場面では物体同士の重なりやすれ違いが頻発し、ID スイッチ数の増加により、正しい追跡ができないという問題がある。

これに対して、長友[2]は複数の小動物を追跡するための複数小動物用 SORT を提案している。小動物は集団で行動し、すれ違いながら高速に移動するため従来の SORT では ID スイッチが発生しやすい。そこで、「滞留」状態を導入して検出対象の重なりによる検出漏れに対応し、「二段階マッチング」の実装により信頼度低下によって追跡が途切れてしまう問題を防止している。これにより、複数小動物追跡時の ID スイッチ数を低減できている。

また、Bodla ら[3]は、物体検出の際に後処理として行われる Non-Maximum Suppression(NMS)に改良を加えることで、検出漏れを防ぐ Soft-NMS を提案

している。NMS は、同一の物体に対して、重複して得られた検出候補の中で、最もスコアの高いもののみ残して、それ以外は除外する処理である。Soft-NMS では除外するのではなく、スコアを減少させて残すことで正しい検出候補を削除してしまう問題を防止している。

本研究では、重なりが大きく検出漏れによる ID スイッチが起きてしまう場面において、ID スイッチ数を削減するために複数小動物用 SORT を元に改良した手法(NasuSORT)を提案する。具体的には、検出漏れを防止するために、検出を行っている YOLO[6]で後処理として行われている NMS 処理を、「Soft-NMS」に変更する。これにより、追跡対象が重なり、信頼度スコアの低下によって削除されていた正しい検出候補を残すことができ、以降の追跡の処理で用いることができる。これまで NMS によって削除されていた誤った検出候補も、SORT で処理されることになるが、複数小動物用 SORT の機能である「二段階マッチング」によって誤った追跡 ID の新規作成を防ぐことができる。この提案手法により、同じ物体に対して複数の追跡 ID が割り振られる誤追跡を防ぎつつ、ID スイッチを削減させる。

以下、2 章では複数物体追跡の従来研究について述べる。3 章では本研究の目的である重なりに強い複数物体追跡を行うための手法を提案する。4 章では提案手法の効果を評価するために行った実験とその結果について述べる。5 章では本研究の結果を要約し、今後の改善点について述べる。

## 2 複数物体追跡の従来研究

### 2.1 複数物体追跡

複数物体追跡とは、動画像中に含まれる人や物などの対象を複数同時に追跡する技術である。複数物体追跡は研究が盛んな分野であり、様々なアルゴリズムが開発されている。複数物体追跡には、物体追跡を行う際に動画像のすべてのフレームをまとめて処理することで高精度を目指すバッチ型と、現在のフレームと以前のフレームのみを使用し高速化を目指すオンライン型の二つがある。しかし、バッチ型ではリアルタイムでの追跡を行うことができず、オンライン型では追跡精度が低いという問題があった。

これに対する改善策として、近年広く使われるようになった「Tracking by Detection(検出による追尾)」というオンライン型の手法がある[4]。Tracking by Detection では、動画の各フレームから追跡対象の物体を検出し、次にフレーム間で同じ物体に対して同一の ID を割り振っていくことで追跡を行う。検出結果と追跡 ID を対応付ける際に、位置や速度パラメータ、外観特徴などを利用して同一の物体であるか識別する。この Tracking by Detection により、オンライン型でも追跡精度を下げず、高速に追跡することができる。具体的な手法として、複数小動物 SORT[2]がある。また、物体検出を行う検出器には YOLO[6]などがある。

### 2.2 複数小動物用 SORT

Simple Online and Real Time Tracking(SORT)[1]は Tracking by Detection のアプローチをとる複数物体追跡手法の一つである。これを元に、家畜の感染症対策や医学の実験で高精度の追跡が行えるように、小動物用の改良を行ったものが、複数小動物用 SORT[2]である。小動物は複数で体を寄せ合いながら行

動するため、オクルージョンが頻発してしまい従来の SORT では追跡が困難であった。これに対応するために、一時的に検出が途切れた場合の対策として、検出が途切れた場所に物体が留まっている「滞留」状態の導入とカルマンフィルタ予測の停止、検出器からの出力結果をすべて利用するように変更している。さらに、誤検出による誤った追跡 ID の新規作成を防止するための二段階マッチングを追加している。複数小動物用 SORT の流れを図 1 に示す。

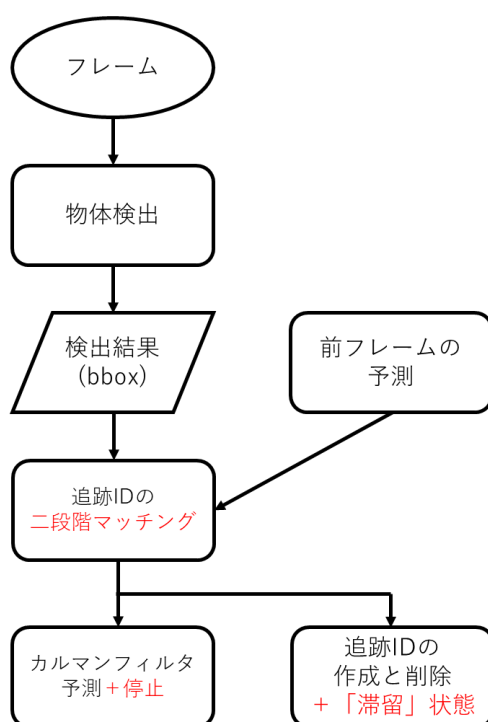


図 1 複数小動物用 SORT の流れ

入力された動画像の各フレームに対して検出器である YOLO を用いて物体検出を適用する。検出結果は物体を囲う矩形であるバウンディングボックス (bbox) と、その矩形が正しく物体を捉えている正確さを表す信頼度スコア(後述)で表現される。検出された bbox に対して、閾値を設け、検出結果を信頼度スコアの高いものと低いものに分ける。それまでの追跡結果(軌跡)から予測された位置(後述)との重なり具合を求め、一番重なりが大きい bbox に対して追

跡 ID を割り当てる。この時、重なり大きさは IoU スコアで表される。図 2 に IoU の式を表す。

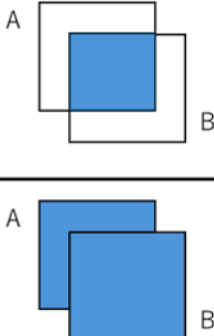
$$\text{IoU} = \frac{\text{AとBの共通部分}}{\text{領域Aと領域Bの和}}$$


図 2 IoU の式

IoU(Intersection of Union)[5]は、二つの領域がどのくらい重なっているかを表す指標で、二つの領域の共通部分を和集合で割ったものである。

保持している追跡 ID と対応付けを行う際に、二段階マッチングを導入している。これにより誤った新規追跡 ID の作成を防いでいる。まず、信頼度スコアが高い bbox からマッチングを行う。その後、一度目のマッチングで残っている軌跡の予測位置と、信頼度の低い検出結果のマッチングを行う。二度のマッチング後に、対応付けられた追跡 ID の新たな位置として新しい bbox の位置を軌跡として更新する。保持している追跡 ID と対応づかなかった bbox は、新規 ID を割り当てられるが、信頼度スコアの高い検出のみを対象とする。bbox と対応づかなかった軌跡は、削除するのではなく、「滞留」状態とする。次フレームの処理のために、追跡対象が等速度で移動していると仮定し、これまでの軌跡を基にカルマンフィルタで次フレームの bbox の位置と大きさを予測する。



## 2.3 YOLO

You Only Look Once(YOLO)[6]は、動画像中から特定の物体を検出する時に使用される代表的なアルゴリズムである。以前の物体検出モデルでは、動画像中から対象物と思われる領域を検出して、それを Deep Neural Network(DNN)の入力として識別処理を行うという二段階で検出を行っていた。しかし、この方法では予測された領域すべてを DNN に入力する必要があるため、処理に時間が掛かるという問題があった。この解決策として、対象物と思われる領域の検出を DNN の一つのレイヤーとし、検出と識別を一貫して行えるようにした end-to-end な物体検出手法が提案された。YOLO も end-to-end な処理によって検出を行っている。

YOLO では、信頼度スコア(Confidence Score)という要素を使用して物体検出を行っている。信頼度スコアは、「それぞれの bbox に物体が入っていて、正確に領域を囲っているかの正確さ」と「各クラスの予測確率(bbox の物体が特定のクラスに属している可能性)」を表している[7]。一連の処理の後、NMS によって確実性の高い bbox のみが残る。

本研究では、YOLO の改良版である YOLOv5[8]を用いる。

## 2.4 NMS

Non-Maximum Suppression(NMS)とは、多くの物体検出手法で共通する後処理である。YOLO などの手法では、bbox 同士の関係については気にせずに推論を行うので、同一の物体に対して重なって検出候補が出力される。この検出候補の中から、信頼度スコアの低いものを除外していき、一つの物体につき一つの bbox が割り当てられるように処理を行っている。

NMS の流れを図 3 に示す。まず、クラスごとにすべての bbox を信頼度スコアの高い順にソートし、一番信頼度スコアが高い bbox 一つが選ばれる。この bbox とクラス内の残った bbox との IoU を求め、重なり具合を判定する。IoU が設定された閾値以上だった場合、信頼度スコアが低い方の bbox を削除する。その後、一番信頼度スコアが高かった bbox 以外でこれを繰り返す。最終的には、重なりが小さかった bbox のみが残る。

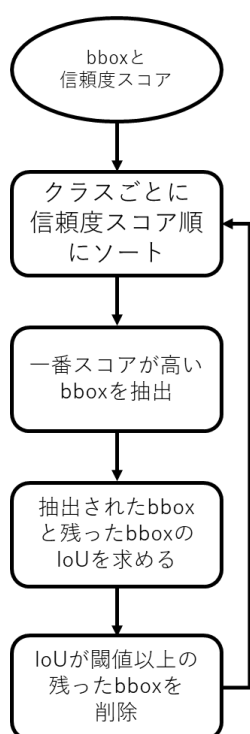


図 3 NMS の流れ

## 2.5 Soft-NMS

従来の NMS は、検出候補として出力された bbox をクラスごとに分けて処理するが、同一クラスの物体が重なって存在している場合に問題が発生してしまう。それは、後方にある物体を検出して出力された bbox の信頼度スコアが十分高くても、手前の物体を検出した bboxの方がより信頼度スコアが高くなり、後方の物体が手前の物体と同一とみなされてしまう事で発生する。これに

より、後方にある物体を正しく検出した検出候補を除外してしまい、検出漏れが発生してしまう。NMS では IoU を使って除外する検出候補を選択するため、重なりが大きいと信頼度スコアが高くても除外してしまう。この問題を解決するために提案されたものが「Soft-NMS」[3]である。Soft-NMS の流れを図4に示す。IoU を求めるまでは従来の NMS と変わらない。IoU が閾値  $k$  以上だった場合、信頼度スコアが低い方の bbox を削除するのではなく、信頼度スコアを減少させて更新し、更新後のスコアが閾値  $s$  以上であれば bbox を残すよう改良されている。

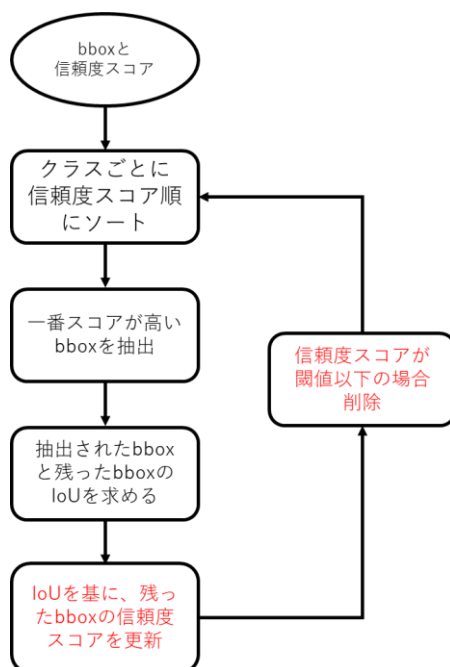


図4 Soft-NMS の流れ

信頼度スコアの更新方法として、二通り提案されている。一つ目は線形的に信頼度スコアを更新する方法(式(1))である。IoU が閾値  $k$  以上であるなら、 $1 - IoU$  の値を元の信頼度スコアに掛けた値とする方法である。

$$f(\text{score}, \text{IoU}) = \begin{cases} \text{score} * (1 - \text{IoU}) & (\text{IoU} \geq k) \\ \text{score} & (\text{otherwise}) \end{cases} \quad (1)$$

もう一つはガウス関数的に信頼度スコアを更新する方法(式(2))で、IoU が大きければより多く信頼度スコアを減少させ、IoU が小さければあまり信頼度スコアを減少させない。

$$f(\text{score}, \text{IoU}) = \text{score} * \exp\left(-\frac{\text{IoU}^2}{\sigma}\right) \quad (2)$$

これらの改良により、重なりによる検出漏れを減少させることができている。加えて、Soft-NMS は既存の NMS に対してほんの少しの変更を加えて実現できるという点で優れている。

## 3 重なりに強い物体追跡

### 3.1 複数物体追跡の問題点

複数物体追跡で近年利用されている Tracking by Detection のアプローチによる代表的な手法として、SORT が挙げられる。SORT はリアルタイムな処理で高速に行えるという利点があるが、追跡対象の重なりや、すれ違いによって ID スイッチが起きてしまうという問題がある。これは検出器によって物体検出を行う時点で、追跡対象の検出ができず、追跡処理を行おうとしても追跡 ID を割り振る bbox が存在せずに追跡が途切れてしまい発生している。

物体検出で物体の重なりによる検出漏れを減らすための手法として、Soft-NMS がある。YOLO 等の物体検出手法の後処理において、NMS の代わりに Soft-NMS を利用することで、物体同士が重なった場合でも検出候補の bbox を残すことができ、検出漏れを減らすことができる。しかし、Soft-NMS を利用した YOLO を SORT の検出処理に利用した場合、同じ物体に対して複数の bbox が残ってしまい、結果として同じ物体に対して複数の追跡 ID が割り振られる誤追跡が起こるといった問題が生じる。

### 3.2 Soft-NMS を追加した複数小動物用 SORT

本研究では、複数物体の同時追跡においてオクルージョンによる ID スイッチを削減するために、複数小動物用 SORT の機能と、検出器の後処理として行われる NMS を改良し検出精度を高めた「Soft-NMS」を組み合わせ、重なりに強い物体追跡手法である、improved SORT with Soft-NMS for robustness against occlusion(NasuSORT)を提案する。具体的には、Soft-NMS の導入で重なりにより削除されていた検出を残し、複数小動物用 SORT の機能で余計に出力された検出から追跡 ID の新規作成を防ぐ。

複数小動物用 SORT では、追跡 ID と新しく検出された bbox の対応付けを行う際に二段階マッチングを行っている。これは信頼度スコアが高い bbox から対応付けを行い、追跡 ID の新規作成も信頼度スコアの高い bbox のみに対して行っている。信頼度スコアが低い bbox は、既に追跡を行っている軌跡に対してのみ対応付けを行う事で誤った追跡 ID の新規作成を防止している。しかし、検出器からの出力結果を使用するので、検出処理の段階で正しく物体を検出できなかった場合、追跡が途切れてしまう問題がある。

Soft-NMS は、検出時にオクルージョンによる検出漏れを防ぐために、重なって検出された bbox を、信頼度スコアの減少という処理を加えて検出結果として残すようにしている。

両者を組み合わせる NasuSORT では、それぞれの問題を解決しつつ、精度向上を目指すことができる。複数小動物用 SORT は検出器として、YOLO を使用しているので、この YOLO の後処理で行われている NMS を Soft-NMS に変更し、検出時点でのオクルージョンによる検出漏れを防ぐ。その際に、物体に対して重複して余計に検出された bbox も出力されてしまう。通常の SORT では余計に検出された bbox に対して、追跡 ID の新規作成を行ってしまい、誤追跡が発生してしまう。しかし、複数小動物用 SORT では、検出された bbox に追跡 ID を割り振る際に二段階マッチングを行うので、Soft-NMS によって信頼度スコアを下げられたが検出されてしまった余計な bbox に対して、追跡 ID の新規作成は行わず、既に追跡を行っている軌跡にのみ対応付けを行う。これにより、誤検出による誤った追跡 ID の新規作成を防止しながら、オクルージョンによって NMS の処理で検出候補の中から削除され、追跡が途切れてしまう問題を改善し、ID スイッチ数を削減することができる。

### 3.3 提案手法の流れ

提案手法の流れを図 5 に示す。まず、入力された動画像のフレームに物体検出を適用する。この時、Soft-NMS を導入した YOLO を用いる。Soft-NMS により通常より多くの検出結果が出力される。閾値を設け、検出結果の高いものと低いものに分ける。この検出結果と保持している追跡 ID の予測位置とで、二段階マッチングを行う。信頼度スコアの高い検出結果からマッチングと追跡 ID の新規作成を行う。その後、Soft-NMS によって信頼度スコアが低下したが、削除されずに残った検出結果とマッチングを行うことで、追跡が途切れることを防ぐ。二度のマッチング終了後、検出結果から追跡 ID の位置を更新する。対応付かなかった軌跡は、滞留状態とする。これらの処理の後、次フレームの処理のために追跡対象の位置を予測する。滞留状態の軌跡については予測を停止する。

複数小動物用 SORT に追加した機能について以下で詳しく説明する。

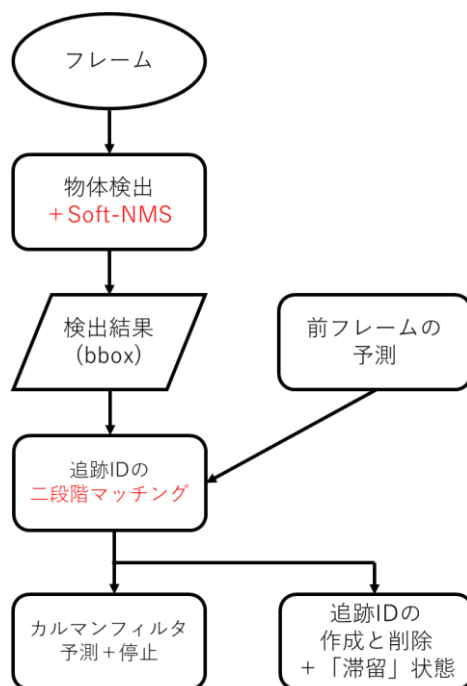


図 5 提案手法の流れ

### 3.3.1 IoU の計算

物体検出を行う YOLO 内部の後処理として行われる NMS を Soft-NMS に変更するにあたって、二つの検出候補の重なり具合を測るために IoU を使用する。その計算を行う機能を追加した。二つの bbox の中心の  $x$  座標と  $y$  座標、幅と高さを受け取り、そこから二つの bbox の面積の和と、重なっている共通部分の面積を求め、共通部分の面積を二つの bbox の和で割り、IoU を求める。

### 3.3.2 Soft-NMS 処理

YOLO の後処理で行われる NMS を Soft-NMS に変更した。まずクラスごとに信頼度スコアの高い順にソートし、一番信頼度スコアが高い検出候補を選ぶ。次に、この検出候補とそれ以外の検出候補との IoU を新しく追加した 3.3.1 節の機能で求める。その値を使って選ばれた検出候補以外の信頼度スコアを更新する。更新した際に信頼度スコアが閾値を下回った検出候補は削除する。その後、選ばれた検出候補以外でソートからまた繰り返す。

### 3.3.3 二段階マッチング

保持している追跡 ID と対応付けを行う際に、二段階マッチングを行う。検出結果を受け取る際に、閾値を設け、信頼度スコアの高いものと低いものに分ける。信頼度スコアが高い bbox の集合からマッチングを行う。この処理が終わると、一度目のマッチングで残っている追跡 ID の予測位置と、信頼度スコアの低い bbox の集合でマッチングを行う。二度のマッチング後に、対応付けられた ID の新たな位置として新しい bbox の位置を軌跡として更新する。保持している ID と対応づかなかった bbox は、新規追跡 ID を割り当てられるが、信頼度スコアの高い bbox の集合のみを対象とする。



## 4 評価実験

### 4.1 実験データ

提案手法の評価のために、動画 1 から動画 4 の 4 本の動画を用意した。各動画の 1 フレームを図 6 に示す。

動画 1 から動画 3 は、改良元である複数小動物用 SORT[2]で評価に使用していた小動物(アフリカヤマネ)の動画である。障害物のない白い箱に 4 匹のアフリカヤマネが入っており、その行動を上方のカメラから 30fps で撮影した物である。各動画のフレーム数を表 1 に示す。動画 1 から動画 3 は、一連の動画から重なりやすれ違いが発生している部分をそれぞれ 40 秒程度の長さに切り取ったものである。

動画 4 は、Fishing Tracking Challenge2024[9]で提供されている Sweat Fish データセットから切り出したアユの動画である。障害物のない白い箱で 10 匹のアユが泳いでおり、その行動を上方のカメラから 30fps で撮影した物である。動画には、各フレームでの各アユの位置(bbox)と追跡 ID が人手により正しく付与されている。フレーム数は表 1 に示した通り、900 枚である。動画 4 は、Sweat Fish データセットで train データとして提供されていた動画の 1 分 30 秒から 30 秒間切り出した動画である。



動画 1



動画 2



動画 3



動画 4

図 6 実験動画の一部

表 1 各動画のフレーム数

	フレーム数
動画 1	1238
動画 2	1632
動画 3	2294
動画 4	900

## 4.2 検出器の学習

追跡を行う際に、物体検出器として検出対象を学習した YOLO を使用する。動画 1~3 のアフリカヤマネの追跡を行う際は、従来研究[2]で使用された学習済みの YOLO を使用する。動画 4 のアユに対しては、物体検出器として YOLO の学習を行った。

YOLO の学習は、物体を映した画像と、その画像内のどの位置に物体が存在するのかを表した位置データが必要になる。画像は、Sweat Fish データセットで train データとして提供されていた動画の最初の 1 分間(1800 フレーム)を 1 フレームごとに画像に分けて使用した。位置データも同じく Sweat Fish データセットで提供されていた位置データを使用した。

学習は 100epoch で行った。学習データに対する最終的な精度は、Precision(適合率)が 0.99734、Recall(再現率)が 0.99083 となっており、検出精度は十分高いと言える。この学習済み YOLO を使って実験を行う。

## 4.3 実験設定

実験を行う際に設定した閾値や、処理について示す。Soft-NMS 処理では、2.5 節で示した式(2)を用いてスコアを減少させた。スコアを減少させる bbox を決める際に使う IoU の閾値  $k$  は 0.45、定数である  $\sigma$  は 0.5、スコアをどの値まで減少させるか決める閾値  $s$  は 0.01 に設定した。二段階マッチングで検出結果を分ける際、スコアが 0.8 より大きい場合はスコアの高い集合とし、0.8 以下の場合はスコアの低い集合とした。

## 4.4 実験結果

実験では、NasuSORT と複数小動物用 SORT、SORT について、ID スイッチの数で比較した。ID スイッチは、追跡中の ID が別の ID に入れ替わってしまうことなので、ID スイッチ数を比較して、追跡に失敗した数を調査できる。

4.1 節で述べた動画 1~4 を目視で調査したそれぞれの追跡結果を表 2~5 に示す。表 2~4 の SORT については[2]で示されていた ID スイッチ数を引用している。複数小動物用 SORT が SORT に比べて ID スイッチ数を削減する事ができた動画 1~3 において、Soft-NMS を導入した NasuSORT の方が ID スイッチ数をさらに減らすことができた。また、10 匹のアユを撮影した動画 4 の場合でも、重なりやすれ違いが発生していたが、複数小動物用 SORT よりも NasuSORT の方が ID スイッチが少ないという結果が得られた。動画 1~4 全てにおいて、NasuSORT の方が ID スイッチが少なかったため、NasuSORT が最も追跡の精度が優れていると言える。

表 2 動画 1 の結果

	ID スイッチ数
SORT[2]	43
複数小動物用 SORT	14
NasuSORT	9

表 3 動画 2 の結果

	ID スイッチ数
SORT[2]	20
複数小動物用 SORT	11
NasuSORT	8

表 4 動画 3 の結果

	ID スイッチ数
SORT[2]	50
複数小動物用 SORT	12
NasuSORT	7

表 5 動画 4 の結果

	ID スイッチ数
複数小動物用 SORT	19
NasuSORT	12

#### 4.5 実験結果の分析

追跡結果から提案手法の分析を行う。図 7 は、Soft-NMS を導入したことによる YOLO の検出数の変化を示している。縦軸は出力された bbox の数、横軸は frame 数を表している。青線は Soft-NMS を導入した YOLO の検出結果で、赤線は従来の YOLO の検出結果である。使用した動画は、4 匹のアフリカヤマネを撮影した動画 1 で、通常の YOLO では 3~5 個の bbox を出力している。それに対し、Soft-NMS を導入した YOLO では三倍ほどの検出結果が出力されている。しかし、従来の YOLO で検出数が減っているフレームでは、Soft-NMS を導入した場合でも、検出数が減少している。

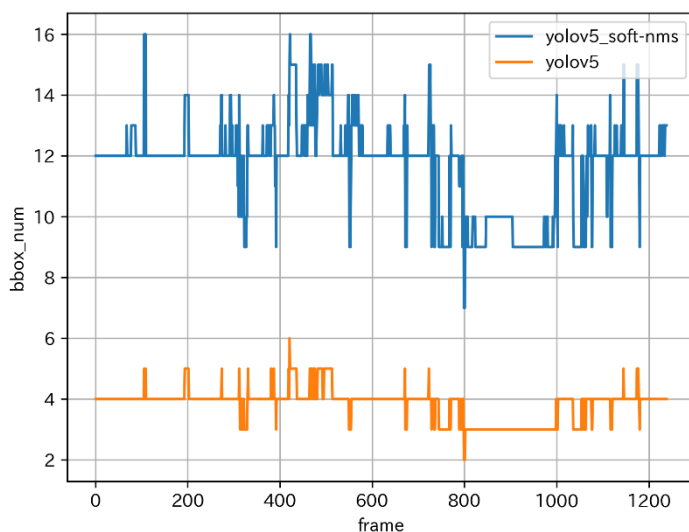


図7 YOLO の検出数

検出結果が増えることで検出漏れを防止することができるが、追跡を行う際に誤検出による誤った追跡 ID の新規作成が起こってしまう。これを二段階マッチングで抑制できているのか確認する。図 8 は、動画 2 を使って様々な条件で追跡を行った際の追跡数と検出数の変化を表している。縦軸と横軸は、図 7 と同じように、検出された bbox の数とフレーム数になっている。青線は、検出を行った YOLO の出力結果を表している。橙色の線は NasuSORT で追跡を行った bbox の数を表しており、緑線は従来の複数小動物用 SORT で追跡を行った bbox の数、赤線は NasuSORT の二段階マッチング機能を削除し、全ての検出結果で追跡 ID の新規作成が可能になっている状態で追跡を行った結果になっている。Soft-NMS の導入によって検出数は増加しているが、NasuSORT は二段階マッチングの効果により追跡数を抑える事ができている。赤線は二段階マッチングの機能が無いため、検出数の増加により間違った追跡 ID が登録されている。この結果から、Soft-NMS の検出結果を二段階マッチングで正しく処理できたことが分かる。また、従来の複数小動物用 SORT で追跡が行えていない 900~1000 フレームの間では、NasuSORT は検出結果の増加により追跡を維持することができている。

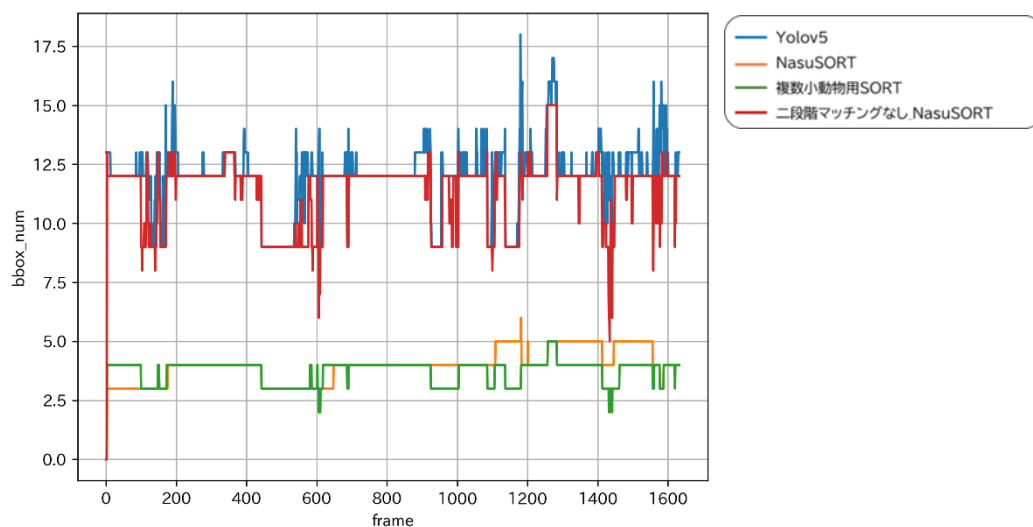


図8 検出数と追跡数の変化

しかし、Soft-NMS を導入した場合でも検出を行えないフレームがあり、追跡漏れが発生している。図9に追跡漏れが発生したフレームを示す。図4には3匹のアフリカヤマネが含まれているが、Soft-NMS を導入したYOLOでも、2つのbboxしか検出できていない。そのため、NasuSORTでも追跡が行えなかった。このように、重なりが大きく、後方にある物体が完全に隠れたり、検出器が検出候補を出力する段階で検出漏れが発生している場合などは、追跡が途切れてしまう。

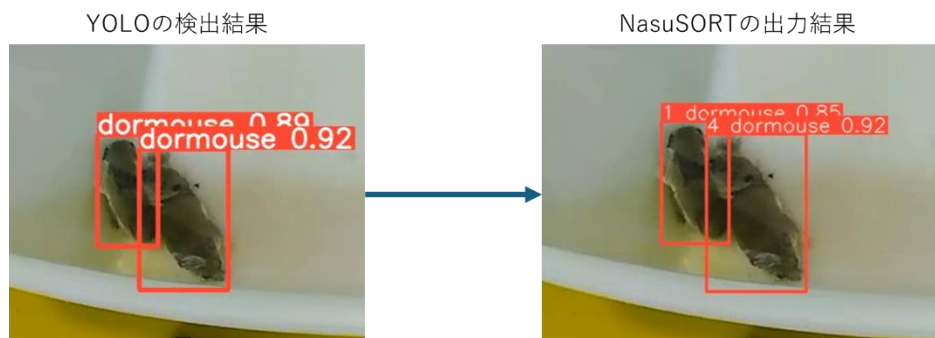


図9 追跡できなかったフレーム

Soft-NMS により検出数は増加したが、二段階マッチングでスコアが高いものは追跡 ID の新規作成、スコアが低いものは既に追跡している ID への関連付けにのみ使い、誤追跡を防止している。しかし、スコアを減少させる際に元のスコアが大きく、IoU が閾値  $k$  に近い場合、減少後もスコアが十分大きくなってしまい、二段階マッチングでスコアが高い集団に振り分けられて、新規追跡 ID を登録してしまっている。図 10 にその状況を示す。図 10 は動画 4 で実験を行った結果で、アユが 10 匹なのに対して、300 フレーム以降は 13~14 個の bbox を出力している。一度追跡 ID が新規作成されると、Soft-NMS によりスコアが減少された検出結果によって軌跡が更新され、誤追跡を続けてしまう問題がある。

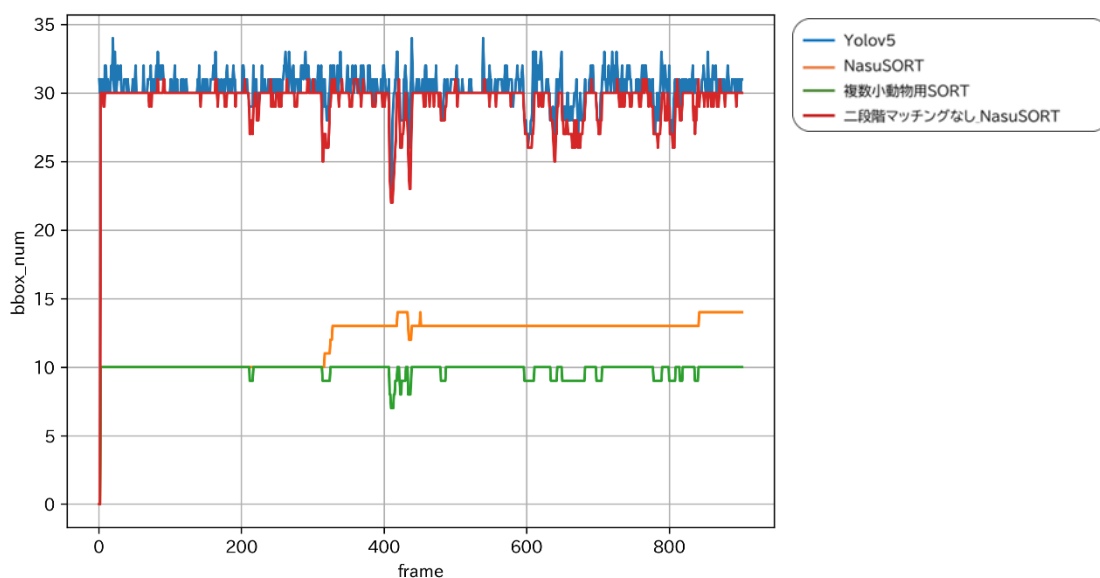


図 10 誤った追跡 ID の新規作成



## 4.6 考察

4つの動画を用いて追跡を行った結果、全ての動画で NasuSORT は複数小動物用 SORT よりも ID スイッチ数が少なくなっていた。さらに、検出数と追跡数を比較すると、Soft-NMS により増加した検出結果による誤った追跡 ID の新規登録を二段階マッチングで防止できていることも確認できた。

この結果から、NasuSORT は、検出漏れを防止する「Soft-NMS」の導入と、複数小動物用 SORT の「二段階マッチング」の機能によって、複数物体追跡を行う際に発生するオクルージョンによる ID スイッチ数を削減したので、追跡精度を高めることができたと考えられる。

## 5 おわりに

本研究では、複数物体追跡においてオクルージョンが発生しても、追跡が途切れないように、複数小動物用 SORT で物体検出を行う YOLO の後処理を NMS から Soft-NMS に変更した NasuSORT を提案した。

NasuSORT では、「Soft-NMS」を導入する事で、重なりが大きく従来の NMS で削除されていた正しい検出結果を、信頼度スコアを減少させた上で残すことができた。また、追跡 ID とのマッチングを行う際も、二段階マッチングにより誤検出による誤った追跡 ID の新規作成も防いでいる。これらの機能により、オクルージョンが発生しても追跡が途切れにくくなり、ID スイッチを削減させている事を確認できた。

今後の課題としては、二段階マッチングを使用しても誤った追跡 ID の新規作成が発生してしまったので、信頼度スコアの減少方法や、追跡 ID のマッチング方法の改良が挙げられる。また、Soft-NMS を導入しても物体を正しく検出できないフレームが存在したので、検出器の改良だけでなく、追跡を行う SORT を、DeepSORT や StrongSORT に変更するなどの改良も必要であると考えられる。

## 謝辞

本論文の作成にあたり、丁寧な指導やアドバイスなど、適切なサポートをして下さった椋木雅之教授には深く感謝いたします。相談のときは、いつも優しくアドバイスを頂きました。本当にありがとうございます。

また、所属する椋木研究室の皆様には様々な助言を頂きました。皆様のおかげで充実した時間を送ることができました。お礼申し上げます。ありがとうございました。

## 参考文献

- [1] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft, 『Simple Online and Realtime Tracking』 2016 IEEE International Conference on Image Processing (ICIP)
- [2]長友 祐磨 『滞留状態と二段階マッチングを追加した SORT による複数小動物の同時追跡』 令和 4 年度 宮崎大学工学部 卒業論文(2022)
- [3] Navaneeth Bodla, Bharat Singh, Rama Chellappa, Larry S. Davis, 『Soft-NMS — Improving Object Detection with One Line of Code』 2017 IEEE International Conference on Computer Vision (ICCV)
- [4] Multiple-object tracking (MOT) アルゴリズム研究の歴史  
<https://tech.acesinc.co.jp/entry/2021/11/08/133336>
- [5] IoU(Intersection over Union): 物体検出における評価指標・ロス関数  
<https://cvml-expertguide.net/terms/dl/object-detection/iou/>
- [6] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi 『You Only Look Once: Unified, Real-Time Object Detection』 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- [7] 【物体検出手法の歴史 : YOLO の紹介】  
[https://qiita.com/cv\\_carnavi/items/68dcda71e90321574a2b](https://qiita.com/cv_carnavi/items/68dcda71e90321574a2b)
- [8] <https://github.com/ultralytics/yolov5>
- [9] Fishing Tracking Challenge2024 <https://ftc-2024.github.io/>