

オートエンコーダーの混合モデルを用いた魚群追跡

芝山 高翔[†] 新里 高行[‡] 波部 齊[§] 川嶋 宏彰[¶]
 兵庫県立大学[†] 筑波大学[‡] 近畿大学[§] 兵庫県立大学[¶]

1 はじめに

生物の複数個体追跡では、各個体の検出だけでなくフレーム間での同一個体の対応付け、すなわち ID の追跡が重要となる。しかし、限られた移動可能範囲の中で多数の個体が動く場合、複数個体追跡時に ID の入れ替わりがしばしば発生する。これは、複数個体が重なることでの検出エラーが主な原因である。個体追跡結果を基に生物の個体間相互作用などの行動分析を行う場合、ID の入れ替わりはノイズになるため望ましくない。

本研究では、水槽内の魚群の追跡を行うにあたり、個体の見た目の類似性が非常に高いという問題設定の下で追跡精度を向上させることを目的とする。このとき、各々が単一個体を生成できるようなオートエンコーダーを複数組み合わせ合わせた混合モデルを用いることで、重なった個体を分離する手法を提案する。

2 関連研究

物体追跡の研究として代表的なものに、SORT (simple online and realtime tracking) [1] が挙げられる。この手法では、物体検出器 (detector) により得られた物体領域を示す bounding box (bbox) と、カルマンフィルタを用いて予測した bbox とのそれぞれの重なりを IoU (intersection over union) により計算する。そして、IoU が最大となる検出個体と予測個体の組み合わせを、ハンガリアンアルゴリズムを用いて探索し、得られた組み合わせをフレーム間で関連付けていく。

生物トラッキングの主な既存ツールとしては、idtracker.ai [2] や TRex [3] がある。これらのツールでは、短いフレーム間で個体の動きを追跡した断片 (fragment) を生成し、各 fragment での個体識別と ID 付与を行ったうえで、fragment を連結していくという手法を採用している。ここで、個体識別には視

覚的情報 (画像特徴) を入力とした畳み込みニューラルネットワークが利用されている。

一方、本研究で扱う魚群の場合、撮影条件によっては各個体の画像が非常に似通っており、画像特徴では個体識別が困難である (図 1(1))。したがって、全個体の見た目が同じであるという問題設定をおいている。そこで提案手法では、個体識別による ID の割り当てではなく、画像を生成可能なニューラルネットワークを用いて重なり状態 (overlap) のモデル化を行うことで、検出エラーに対する頑健性を高めていくという新たなアプローチを採る。

3 提案手法

提案手法の流れは以下の通りである (図 1)。個体追跡を行う動画中の時刻 t のフレームを F_t とする。また、フレーム内に存在する個体の総数は一定であると仮定する。まず、画像の領域分割 (segmentation) により対象の領域 (segment) 集合 S_t とその bbox 集合 B_t を得る。この段階では 1 segment に複数個体が含まれる場合もある。そこで、 S_t の中で閾値以上の面積を持つ k 番目の segment の bbox を $B_{t,k}$ とし、その segment から overlap 画像 $O_{t,k}$ を作成する。

一方、tracker 側では F_{t-1} の追跡情報 (カルマンフィルタの状態) を基に、 F_t での各個体の bbox を予測した集合 B'_t を計算し、これより $B_{t,k}$ 内に存在すると考えられる予測個体数 $n_{o,k}$ を計算する。

複数個体が overlap した画像 $O_{t,k}$ に対し、オートエンコーダーの混合モデルを利用して個体の分離を行う。具体的には、 $n_{o,k}$ 個の (潜在) 変数によるデコード結果 (生成画像) の混合画像 $G_{t,k}$ が最も $O_{t,k}$ と類似するように、各変数を最適化する。すると、各個体の分離画像が得られるため、分離後の新たな bbox 集合 $B''_{t,k}$ に基づいて B_t を修正し、tracker による ID の割り当てを実行する。

以下、個体分離、segmentation と overlap 判定、および $n_{o,k}$ の計算について具体的方法を説明する。

3.1 オートエンコーダーの混合モデル

混合モデルを構成する $n_{o,k}$ 個の生成モデルには、同一の重みパラメータを持つ畳み込みオートエンコーダー (Convolutional Autoencoder, CAE) を用いる。

Fish School Tracking with a Mixture of Auto-encoders

[†] Takato Shibayama, University of Hyogo

[‡] Takayuki Niizato, University of Tsukuba

[§] Hitoshi Habe, Kindai University

[¶] Hiroaki Kawashima, University of Hyogo

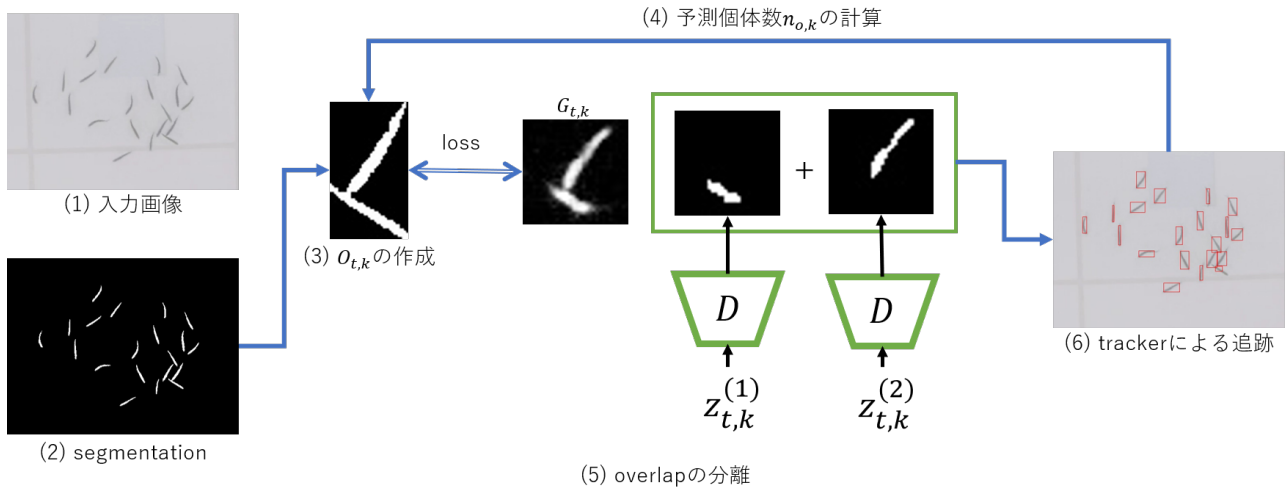


図1 提案手法の流れ. 混合画像 $G_{t,k}$ が overlap 画像 $O_{t,k}$ と最も近くなるように各変数 $z_{t,k}^{(i)}$ を最適化する.

入出力画像の解像度はいずれも 32×32 , 入力画像は、魚領域が白、背景が黒の二値画像とする. 入力画像はエンコーダー E により $10 \times 2 \times 2$ 次元の潜在変数 z に圧縮され, z はデコーダー D で再び二値画像へ復元される. CAE のネットワークは, overlap していない個体の二値画像を学習データとして, 1 個体の画像を再構成できるように予め学習しておく.

$O_{t,k}$ の分離は, デコーダー D に入力する $n_{o,k}$ 個の変数 $z_{t,k}^{(i)}$ ($i = 1, \dots, n_{o,k}$) の初期値を決め, 各変数より生成される画像 $D(z_{t,k}^{(i)})$ を混合した画像 $G_{t,k}$ と $O_{t,k}$ との平均二乗誤差損失 (loss) に基づいて, 指定した回数まで $z_{t,k}^{(i)}$ の更新を繰り返していく.

3.2 Segmentation と overlap 判定

Segmentation では, F_t と背景画像による背景差分と閾値処理により二値化を行い, 連結成分のラベリングで各個体を取り出す. 用いる背景画像は, 動画のフレームを一定間隔で抽出し各画素でその中央値を求めることで生成する. B_t は S_t の x 座標, y 座標の最小値と最大値から計算する. Overlap の判定は, segment の面積が, ある基準値から一定の倍率を超えるかどうかに基づいて行う. この基準値には過去 3 フレームの segment 面積の中央値を用いる.

3.3 予測個体数 $n_{o,k}$ の計算

まず, B'_t と $B_{t,k}$ の IoU を計算する. この IoU が閾値以上となる B'_t の bbox が実際の F_t において $B_{t,k}$ の範囲内に存在する個体と考えられるため, この個体を $B_{t,k}$ における予測個体数 $n_{o,k}$ とする.

4 評価実験

提案手法による 2 個体の分離例を図 2 に示す. 各個体画像の復元は十分ではないものの, 2 個体が正

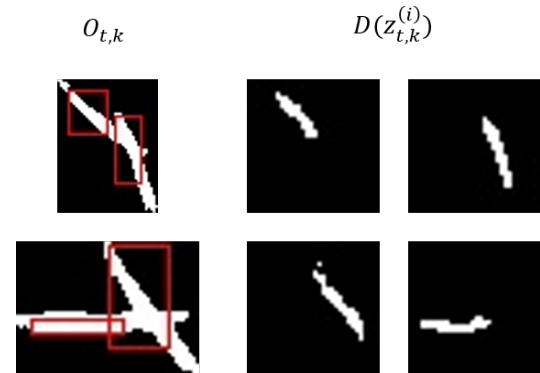


図2 分離結果例 (2 例). 左: 入力画像および分離後の bbox 集合 $B'_{t,k}$ (赤枠). 右: 最適化後の変数 $z_{t,k}^{(i)}$ ($i = 1, 2$) より復元された分離画像.

しく分離されていることが分かる. 本手法を 2 節で述べた SORT などの個体追跡に組み込んだときの, 個体 ID の入れ替わりに関する精度評価を今後の課題とする.

謝辞 本研究の一部は科研費 JP21H05302 の補助を受けて行った.

参考文献

- [1] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3464–3468, 2016.
- [2] Francisco Romero-Ferrero, Mattia G. Bergomi, Robert C. Hinz, Francisco J. H. Heras, and Gonzalo G. de Polavieja. idtracker.ai: tracking all individuals in small or large collectives of unmarked animals. *Nat Methods*, Vol. 16, pp. 179–182, 2019.
- [3] Tristan Walter and Iain D Couzin. Trex, a fast multi-animal tracking system with markerless identification, and 2d estimation of posture and visual fields. *eLife*, Vol. 10, p. e64000, feb 2021.