

生体蛍光観察動画像の深度を考慮した 深層学習による細胞追跡精度の改善

嶋田 彩人¹ 瀬尾 茂人^{1,a)} 繁田 浩功¹ 間下 以大^{1,2} 内田 穰³ 石井 優^{3,4} 松田 秀雄¹

受付日 2019年1月31日, 再受付日 2019年3月20日,
採録日 2019年3月31日

概要: 二光子励起顕微鏡や蛍光タンパク質等の技術発展により, 生体内の細胞をリアルタイムで観察可能となっている. 細胞動態の解析を行うためには, まず細胞の動きを自動で追跡することが基本的なタスクである. しかし, 観察する対象や用いる技術に依存して画像の見た目や細胞の特徴や性質が大きく変化するため, 追跡のためのアルゴリズムや特徴量を個別に設計する必要があった. この問題に対して, 追跡対象の物体の特徴を随時学習しながら追跡を行う教師あり学習による手法が提案されており, Action-Decision Network (ADNet) は畳み込みニューラルネットワークと強化学習を組み合わせることで物体追跡を行う方法の1つである. 本研究では, 転移学習により ADNet を蛍光生体イメージングで得られた動画像での細胞追跡に応用する方法を提案する. また, 3次元の細胞動画像の深度の情報を疑似的な深度マップとして入力に用いることで, 細胞の追跡精度が向上することを示す.

キーワード: 細胞画像処理, 物体追跡, 深層学習, 蛍光生体イメージング, 畳み込みニューラルネットワーク, 強化学習, 最大値投影法, 深度マップ

Cell Tracking based on Deep Learning with Depth Information for Intravital Fluorescence Imaging Data

AYATO SHIMADA¹ SHIGETO SENO^{1,a)} HIRONORI SHIGETA¹ TOMOHIRO MASHITA^{1,2} YUTAKA UCHIDA³
MASARU ISHII^{3,4} HIDEO MATSUDA¹

Received: January 31, 2019, Revised: March 20, 2019,
Accepted: March 31, 2019

Abstract: Advances in technology such as two-photon excitation microscopy and fluorescent proteins have enabled the observation of intravital cells. In order to analyze cell dynamics, it is a fundamental task to automatically track cell movement. However, with the diversity in the appearance of the image and the characteristics of cells, the tracking algorithm must be designed individually. To address this challenge, supervised learning approaches have been developed. Action-Decision Network (ADNet) is one of the methods to track the object by combining convolutional neural network and reinforcement learning. In this study, we propose the use of ADNet by transfer learning for cell tracking in the image sequence obtained by intravital fluorescent imaging. Moreover, we show that depth information of the cells in three-dimensional image sequences improves cell-tracking accuracy.

Keywords: bioimage informatics, object tracking, deep learning, intravital fluorescence imaging, reinforcement learning, maximum intensity projection, depth map

¹ 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology,
Osaka University, Suita, Osaka 565-0871, Japan
² 大阪大学サイバーメディアセンター
Cybermedia Center, Osaka University, Toyonaka, Osaka
560-0043, Japan
³ 大阪大学大学院医学系研究科
Graduate School of Medicine, Osaka University, Suita, Osaka
565-0871, Japan

1. はじめに

近年の二光子励起顕微鏡を始めとした光学観察技術の発

⁴ 大阪大学大学院生命機能研究科
Graduate School of Frontier Biosciences, Osaka University,
Suita, Osaka 565-0871, Japan
a) senoo@ist.osaka-u.ac.jp

展により、多種多様かつ大量の細胞動画画が日々産出されている。たとえば、二光子励起顕微鏡で生きたマウスの骨組織内を観察することにより、骨代謝における破骨細胞と骨芽細胞の分布と活動のバランスの変化する様子が [1]、また皮膚を観察することにより、顆粒球等の免疫細胞の動きが急性炎症やアレルギーといった条件によって変化する様を観察することができる [2]。このような動画画の中から細胞の動きを抽出した結果を解析することで、薬効の定量的な評価や病態の解明に貢献することが期待される。一方で、バイオイメージングのハイスループット化が進む現在、手動による細胞の追跡は負担が大きく、また観測者の主観によるバイアスも危惧されるため、画像処理による自動追跡技術の開発は喫緊の課題となっている。

二光子励起顕微鏡は、高い組織透過性と低い組織侵襲性により、生体内の観察に適した方法である。蛍光の励起は焦点面でのみ起こるため Z 軸方向に高い空間解像度を持ち、焦点を Z 方向に移動しながら断層像の繰り返し撮影（スタック）することで、対象を 3 次元的に観察することができる。加えて一定の時間間隔をおいて撮影を続けることにより 3 次元の動画画を得ることができ、これをタイムラプスという。ただし、3 次元でのタイムラプスの場合は、時間的な制約が存在し、Z 方向へのスタックの枚数とタイムラプスの時間間隔は反比例することになる。特に免疫細胞のような移動する細胞の動態を観察するような場合では、細胞が大きく動いてしまう前に次フレームの撮影をはじめなければいけないため、Z スタックの枚数は必然的に少なくなることになる。よってこのような場合には、そもそも Z 方向に広く観察することは行わず、比較的狭い領域の Z スタックを撮影し、最大輝度投影法（Maximum Intensity Projection, MIP）処理を行って 2 次元に投影した動画を解析対象とすることが行われる。MIP によって Z 方向の情報は失われるものの、画像としては S/N 比が向上し、細胞の検出が行いやすいという利点がある。図 1 の左側は、ある時点において Z 方向にスタックした断層像の例であり、右側はその MIP 画像である。この例では、緑色で見えているのが追跡を行う対象となる細胞（この画像では主に顆粒球）であり、青色で見えているのは皮膚の線維、いわば背景である。

このような蛍光生体イメージングにおいても、観察する対象は多岐にわたるため、個別に適したアルゴリズムを設計・調整するのは困難である。そこで本研究では、一般物体追跡のために開発された深層強化学習に基づく方法である ADNet [3] を、蛍光生体イメージング画像における細胞追跡へと応用する方法を提案する。これは初期値として与えた対象の特徴を随時更新しながら追跡を行う方法であり、細胞の特徴を個別に設計する必要がないのがメリットである。本研究では、一般物体追跡のデータセットによってニューラルネットワークを事前学習し、これを皮膚内の

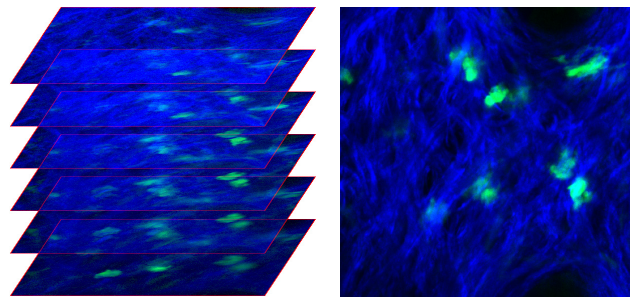


図 1 Z スタックと MIP 画像

Fig. 1 Z stack and MIP image.

細胞の遊走を観察したデータに転移学習することで、細胞の追跡が可能であることを示す。また、蛍光生体イメージングにおいては、いずれの細胞も単色の輝点の集合として観察されることから、それぞれの細胞を見た目で区別することが難しいという問題もある。このアピランス情報の乏しさを補うために、本研究では MIP の際の深度の情報を疑似的な深度マップとして入力に加えることで、細胞の追跡精度の向上を試みる。

以下 2 章で関連研究、3 章で深層強化学習による物体追跡の従来手法である Action-Decision Network (ADNet) について述べ、4 章に生体蛍光観察動画の深度を考慮した深層学習による細胞追跡精度の改善手法を示す。5 章では実際に二光子励起顕微鏡による観察結果である動画に対して提案手法を適用し細胞追跡の精度評価を行い本研究の有用性を示す。

2. 関連研究

細胞の追跡の自動化は、個別に考案・調整されたアルゴリズムの開発により、各論的に解決されてきた。多くの場合において基本的な細胞追跡の方法は、まず各フレームにおいて細胞を検出し、同一である可能性が高い細胞を対応付けるというものである。しかし、バイオイメージングのデータは、一般に低コントラストでノイズの大きい動画画になることが多いため、細胞を正確に検出すること自体が難しい問題である。最も単純な細胞検出の方法は閾値処理であるが、かなりの好条件が揃わなければ十分な精度を得ることはできない。近年では教師あり学習を利用する方法も多く開発されている。判別器として SVM (Support Vector Machine) [4] を用いるものから、畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) を利用する方法も多く提案されており、広く利用可能なソフトウェアとしても公開されている [5], [6]。

各フレームにおいて検出した細胞を対応付けるのに、最も単純な方法は最近隣 (Nearest neighbor) 法である。これは時間の前後の画像において最も距離が近いものを同一の対象であるとする方法であるが、密集状況下では精度が得られない。そのため、全フレーム間において最適な対応

付けを計算する方法も提案されている [7]. その他, Mean shift 法のように対象の色ヒストグラムを特徴量として反復法で追跡を行う方法 [8] や, パーティクルフィルタ [9] 等が, 追跡する細胞の特徴に応じて個別に調整され, 利用されている.

一方で, 一般物体追跡 (Visual Object Tracking) は精力的に研究が行われている分野である. これは, 動画開始時に指定した追跡対象を, 対象や背景の変化に適応し続けながら追跡するタスクであり, VoT Challenge [10] のようなコンペティションも毎年行われている. 物体追跡における真値とともに公開されるデータセットは, 様々な方法の開発や評価に利用されている. 特に, 近年では畳み込みニューラルネットワークや再帰型ニューラルネットワーク (Recurrent Neural Network) 等, 深層学習を組み合わせた方法による精度の向上が著しい [11].

CNN を用いた一般物体追跡手法は多く提案されているが, それらの多くは “Tracking-by-Detection” と呼ばれるタイプの手法である. これは, 追跡対象周辺に候補となる矩形領域を大量に生成し, それらすべてについて追跡対象であるか否かを分類するような追跡器を教師あり学習を用いて訓練するというものであるが, いくつかの問題点を抱えている. “Tracking-by-Detection” では追跡時に, 物体近傍の矩形領域について分類を行う必要があるため, 非効率的な追跡となってしまう. また, 訓練に用いる動画画像中の全フレームに対して真値となる矩形領域の座標が必要となるため, 長時間の動画画像を用いての訓練が難しいという問題点が存在する. これに対して, “Action-Driven” な追跡手法は, 従来の “Tracking-by-Detection” な追跡手法とは異なり, 強化学習アルゴリズムをベースとしている. 次節では, “Action-Driven” な一般物体追跡手法である Action-Decision Network (ADNet) [3] についての説明を行う.

3. ADNet を用いた一般物体追跡

3.1 問題設定

ADNet を用いた際の物体追跡問題は, マルコフ決定過程に従う. マルコフ決定過程は, ステップ t における行動 $a_t \in \mathcal{A}$, 状態 $s_t \in \mathcal{S}$, 状態遷移関数 $s_{t+1} = f(s_t, a_t)$, 報酬関数 $r(s_t)$ の 4 要素によって定義される.

図 2 は ADNet の追跡および, 追跡中の処理の流れを表したものである. ADNet では, 物体の位置を矩形の操作によって追跡する. 状態 (矩形の位置/行動履歴) が方策に入力され, 行動 (矩形の変形/移動) が決定される. 決定された行動をもとに矩形が移動され, 行動履歴が更新されることによって次の状態へと遷移する. この一連の流れをステップと呼び, 1 フレーム中に数回のステップが実行される. 行動を決定するための方策は教師あり学習と強化学習によって訓練される.

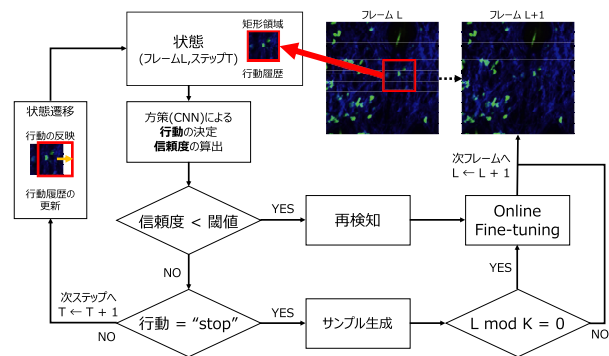


図 2 ADNet による追跡の流れ

Fig. 2 Pipeline of the tracking using ADNet.



図 3 11 種類の行動

Fig. 3 11 Actions.

方策の学習については 3.2 節で説明を行い, 以下では ADNet を用いた物体追跡問題における各要素の定義について説明を行う.

3.1.1 行動

行動空間 \mathcal{A} は以下の 11 種類の行動によって構成される.

- 矩形を上下左右に移動させる (4 種類)
- 矩形を上下左右に 2 回移動させる (4 種類)
- 矩形を拡大/縮小する (2 種類)
- 現在フレームにおける矩形の移動を終了する (1 種類)

それぞれの行動は 11 次元の one-hot 表現の形で扱われる. また, 図 3 は 11 種類の行動の模式図である.

3.1.2 状態

状態 s_t は (p_t, d_t) の組で定義される. ここで, p_t はステップ t における画像中の矩形領域を表しており, d_t はステップ t における過去 k 回の行動を表すベクトルである. 矩形 p_t は, 4 次元ベクトル $[x_t, y_t, w_t, h_t]$ で表現される. ここでは画像左上を原点とした座標系を用いており, (x_t, y_t) はステップ t における矩形左上の座標を, (w_t, h_t) はステップ t における矩形の幅, 高さをそれぞれ表している. 行動履歴 d_t における, それぞれの行動 a_{t-i} ($i = 1, 2, \dots, k$) は one-hot 表現として保存される.

3.1.3 状態遷移

ステップ t において, 状態 s_t のもとで行動 a_t が選択された場合, 状態遷移関数 $f(s_t, a_t)$ に従いステップ $t+1$ における状態 s_{t+1} が決定される. 矩形の移動に関する状態遷移 $p_{t+1} = f_p(p_t, a_t)$ はそれぞれ次式で表される.

$$x_{t+1} = x_t \pm \alpha w_t \tag{1}$$

$$y_{t+1} = y_t \pm \alpha h_t \tag{2}$$

ここで, α は移動倍率を表すパラメータである.

また, 矩形 p_t の拡大縮小が選択されたときの状態遷移は次式で表される.

$$x_{t+1} = x_t \pm \frac{\beta - 1}{2} w_t \quad (3)$$

$$y_{t+1} = y_t \pm \frac{\beta - 1}{2} h_t \quad (4)$$

$$w_{t+1} = \beta w_t \quad (5)$$

$$h_{t+1} = \beta h_t \quad (6)$$

β は拡大倍率を表しており、拡大が選択された場合は加算を、縮小が選択された場合は減算を行う。「終了」の行動が選択された場合は、現在フレームの状態遷移は終了し、報酬を受け取り次フレームの初期状態へと遷移する。また、状態遷移において、矩形の移動と同時に行動履歴 d_t の更新が $d_{t+1} = f_d(d_t, a_t)$ に従って行われる。行動履歴は k ステップの行動を保存するので、最も古い行動である a_{t-k} が破棄され、行動 a_t が新たに保存される。

3.1.4 報酬

あるフレーム l における真値 g_l が存在する場合、報酬関数は以下のように定義される。フレーム l における最終ステップ、つまり「終了」の行動が選択されたステップを $T_l = (t | a_{t,l} = \text{“終了”})$ としたとき、

$$r(s_{t,l}) = \begin{cases} 0 & (t \neq T_l) \\ 1 & (IoU(p_{T_l}, g_l) > 0.7) \\ -1 & (IoU(p_{T_l}, g_l) \leq 0.7) \end{cases} \quad (7)$$

IoU は Intersection-over-Union(IoU) を表しており、 $IoU(p_a, p_b) = \frac{|p_a \cap p_b|}{|p_a \cup p_b|}$ で定義される。 $|p_a \cap p_b|$, $|p_a \cup p_b|$ はそれぞれ2つの矩形領域 p_a, p_b の積集合部分のピクセル数、和集合部分のピクセル数を表している。ここで、ADNet を用いた物体追跡問題における報酬関数は、行動 a_t や遷移後の状態 s_{t+1} には依存しない。また、真値 g_l が存在しない場合は、「終了」の行動が選択された場合においても報酬 0 が与えられる。

3.2 ネットワークの学習

3.2.1 教師あり学習のためのラベル生成

教師あり学習において、ADNet は出力を 2 つ持つマルチタスク学習を行う。与えられた矩形領域に対して、一方では最適な「行動」を選択するようなタスクを学習し、もう一方では追跡対象である確率を表す「信頼度」を出力するタスクを学習する。そのためには、追跡対象物体を含む矩形（正のクラス）と含まない矩形（負のクラス）を多数生成し、クラスラベルをつける必要がある。また、それらの矩形に対して適切な行動ラベルをつける必要がある。

物体追跡問題における真値は矩形領域の座標として与えられ、それらを基に行動ラベルおよびクラスラベルを生成する。真値の矩形 g から座標やスケールを変化させた N 個の矩形 $\{p^{(i)}\}_{i=1}^N$ を生成する。生成された矩形に対して、3.1.1 項で定義したすべての行動を用いて矩形の移動を行う。その中で最も IoU が高かった行動を、矩形に対応する

行動ラベル $a^{(i)}$ として扱う。

$$a^{(i)} = \arg \max_{a \in \mathcal{A}} IoU(f_p(p^{(i)}, a), g) \quad (8)$$

ここで $f_p(p, a)$ は矩形の移動に関する状態遷移関数を表している。さらに、生成された矩形 $p^{(i)}$ に対して、真値である矩形 p と一定より高い IoU を持つ場合は正のクラス、そうでない場合は負のクラスとして扱う。

$$y^{(i)} = \begin{cases} 1 & (IoU(p^{(i)}, g) > 0.7) \\ 0 & (\text{otherwise}) \end{cases} \quad (9)$$

$y^{(i)}$ は矩形領域 $p^{(i)}$ に対応するラベルを表している。これらによって、真値となる矩形領域から、矩形領域 $p^{(i)}$ 、行動ラベル $a^{(i)}$ 、クラスラベル $y^{(i)}$ を組として持つサンプル $(p^{(i)}, a^{(i)}, y^{(i)})$ が生成できる。

3.2.2 教師あり学習によるパラメータ更新

ニューラルネットワークは、3.2.1 項において説明した 2 種類のタスクについて、マルチタスク学習を行う。訓練に使用されるミニバッチは毎エポック、それぞれの動画像からサンプル (p_i, a_i, y_i) を N 個ランダムに抽出し作成する。また、損失関数は以下のように定義されており、確率的勾配降下法を用いて最小化を行う。

$$L = \frac{1}{N} \sum_{i=1}^N L_1(a_i, \hat{a}_i) + \frac{1}{N} \sum_{i=1}^N L_2(y_i, \hat{y}_i) \quad (10)$$

ここで、 L_1, L_2 はクロスエントロピー関数を、 a_i, y_i はそれぞれ行動/クラスのラベルを、 \hat{a}_i, \hat{y}_i はそれぞれニューラルネットワークによって予測された行動/クラスを表している。

3.2.3 強化学習によるパラメータ更新

3.2.2 項で説明した方法によって教師あり学習によるニューラルネットワークの訓練を行った後、ネットワークをパラメータ化された方策として扱い、このパラメータを強化学習により更新していくことで方策を改善する。訓練済みのネットワークのパラメータを初期値に方策ネットワークとして強化学習を行う。強化学習のための事前学習として教師あり学習を行っておくことで、訓練が効率的に進行する可能性があることが確認されている [12]。ここでは現在の方策を用いて追跡を行い、その結果によって報酬を受け取りパラメータを更新する。強化学習の目的は方策ネットワークの改善であるため、教師あり学習において訓練を行った信頼度を出力するタスクについては、強化学習において訓練は行わない。

強化学習におけるパラメータ更新および追跡シミュレーションの手順について、ここでは詳細な記述を行う。まず、シミュレーションに用いる動画像について、 L フレームのシーケンス $\{V_l\}_{l=1}^L$ およびそれらの真値 $\{G_l\}_{l=1}^L$ をランダムに抽出する。このとき、シーケンス中の全フレームに真値を持っている必要はなく、最終フレームにのみ最低限真

値を持っていけばよい。抽出したシーケンスに対して現在の方策ネットワークを用いて、追跡を行う。シーケンスにおける l フレーム目のステップ t における状態を $s_{t,l}$ としたとき、これを方策となるニューラルネットワークに入力することで、行動 $a_{t,l}$ を獲得する。このとき、以下の式に従って行動が選択される。

$$a_{t,l} = \arg \max_{a \in \mathcal{A}} p(a|s_{t,l}; W) \quad (11)$$

ここで、 $p(a|s)$ はニューラルネットワークの出力となる確率分布を、 W はニューラルネットワークにおける現在のパラメータを表している。状態 $s_{t,l}$ と行動 $a_{t,l}$ が決定した後、状態遷移関数によってステップ $t+1$ の状態 $s_{t+1,l}$ が決定される。また、各ステップの行動後には報酬が与えられるが 3.1.4 項に従い、「終了」の行動が選択された場合は 1 もしくは -1 の報酬が与えられ、それ以外の場合は 0 が与えられる。

1 つのシーケンスの追跡が終了後、与えられた報酬を用いて方策となるニューラルネットワークのパラメータ W の更新を行う。パラメータ更新は確率的勾配上昇法 [13] によって行われ、勾配は以下の式で計算される。

$$\Delta W = \frac{1}{L} \sum_{l=1}^L \frac{1}{T} \sum_{t=1}^T \frac{\partial \log p(a_{t,l}|s_{t,l}; W)}{\partial W} \rho_{t,l} \quad (12)$$

ここで、 T はステップ数を表しており、 $\rho_{t,l}$ は以下の式で定義される。これは、フレーム l の最終ステップ T_l における報酬を表している。

$$\rho_{t,l} = r(s_{T,l}) \quad (13)$$

3.2.4 オンラインアダプテーション

一般に、訓練に用いた動画と、追跡を行う動画は異なることが多い。そのため、実際に追跡を行うときに、より追跡を行う動画のドメインにネットワークを適応させることが望ましい。そこで追跡時に状況に応じてサンプル生成や Fine-tuning、および物体の再検知を行う手法 [14] が提案されており、ADNet においてもこの処理が追跡中に行われている。この手法を本論文では、オンラインアダプテーションと呼ぶ。

図 2 下部は、追跡中に行われるオンラインアダプテーションの流れを表している。あるフレーム L 、ステップ T において、与えられた状態から方策である CNN に従って、行動と信頼度が算出される。ここで、信頼度の値が与えられた閾値未満であった場合は、追跡対象物体を見失ったという判定がなされる。このとき、図 2 中の「再検知」処理が行われる。再検知処理では、現在の矩形領域から座標やスケールを微小に変化させた矩形領域を多数生成し、それらを CNN に入力することで信頼度を計算する。その中から信頼度について上位 5 個の矩形を選択し、それらの座標

の平均を新しい矩形領域として扱う。一方で、初期フレームおよび物体を見失わずに「終了」の行動が選択されたフレームでは、追跡中の動画を基に Fine-tuning を行うためのサンプル生成が行われる。サンプル生成は、3.2.1 項で述べた方法と同様にして行われる。初期フレームを含む一定間隔のフレームで、あるいは再検知が行われたフレームで、Fine-tuning が実行される。Fine-tuning では、追跡中に蓄積されたサンプルを用いて 3.2.2 項で述べた教師あり学習と同様にして随時ネットワークの訓練を行うことで、より追跡中の動画に適したネットワークに更新する。

4. 生体蛍光観察動画の深度を考慮した細胞追跡精度の改善

4.1 概要

本手法では、一般物体データセットによって訓練された ADNet を拡張する。教師あり学習および強化学習を用いて細胞動画データセットの転移学習をすることで、3次元蛍光動画中の細胞追跡を行う。ネットワークの訓練および物体追跡の流れは、3章で示した手順に従う。3次元蛍光動画を MIP によって 2次元画像に投影して解析するという特性上、Z 軸の情報が失われてしまい、細胞が密集しているような状況下では追跡は困難になる。この問題を解決するために、MIP を行う際に Z 軸方向の情報を擬似的な深度として保持し、ネットワークの入力に用いる。これによって MIP 処理された蛍光動画を扱いながら、本来であれば失われる Z 軸方向の深度情報を考慮した訓練を行うことができ、より正確な細胞追跡が可能になる。

4.2 深度成分の入力

MIP によって、Z 軸方向の情報は失われてしまう。そこで、MIP において投影される最大輝度値を持つ z 座標を保存しておき、それを物体追跡のための入力として加えることで、Z 軸方向の情報を考慮した追跡を行う。ここでは、入力として扱う z 座標の情報を疑似深度と呼ぶ。

疑似深度情報は、追跡対象となる細胞を表す蛍光のチャネルから計算する。疑似深度画像是以下の式によって求められる。

$$D(x, y) = \arg \max_z I(x, y, z = k) \quad (k = 1, 2, \dots, Z) \quad (14)$$

ここで、 $D(x, y)$ は疑似深度画像の座標 (x, y) における値を、 $I(x, y, z)$ は対象とするチャネルの 3次元画像における座標 (x, y, z) の輝度値を表している。このとき、生成される疑似深度画像のノイズを低減するために、あらかじめ 3次元画像の各スライスに中央値フィルタによる平滑化を行う。また、本手法は 2次元動画中の物体追跡のためにも 3次元動画同様に用いることができる。その場合は疑似深度を画像中の全画素について 0 として扱う。

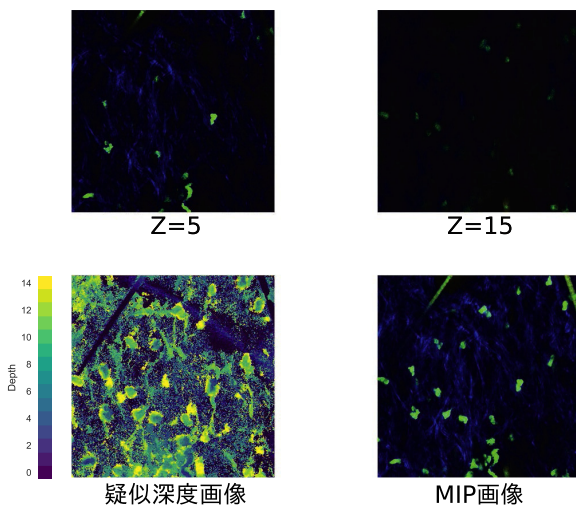


図 4 疑似深度画像の生成例

Fig. 4 Example of pseudo-depth image.

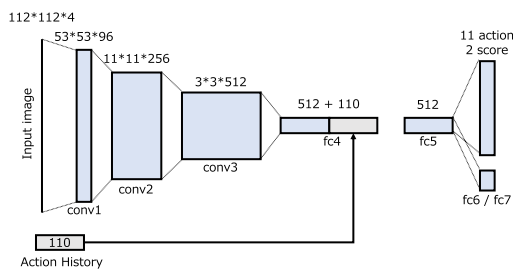


図 5 ネットワーク構造

Fig. 5 Network Structure.

図 4 は、疑似深度画像の生成例である。Z 軸方向に 15 枚スタックされている画像に対して MIP を適用したものである。上 2 つの画像はそれぞれ元画像から $z = 5$, $z = 15$ を抜き出したものを、左下の画像は疑似深度画像を、右下の画像は MIP 画像を表している。

4.3 ネットワーク構造

図 5 は、本手法で使用した CNN の構造を示している。これは、図 2 の方策に該当する。ネットワーク構造は VGG-M モデル [15] を用いている。ネットワークは全体で 3 層の畳み込み層 (conv) と 4 層の全結合層 (fc)、および 2 層のプーリング層からなる。プーリング層は、conv1 層および conv2 層の後に含まれており、最大値プーリングによる処理が行われる。fc4 層および fc5 層の直後には Dropout 層が使用されている。活性化関数には出力層ではソフトマックス関数を、それ以外の層では ReLU を用いている。畳み込み層の各パラメータは、ImageNet [16] によって訓練されたパラメータによって初期化されている。また、画像の入力部分は疑似深度を含むため、本手法では 4 チャンネルとなっている。ネットワークに入力される画像は矩形を 112×112 の大きさにリサイズしたものである。行動履歴の入力部分は、11 行動の 10 ステップからなる 110 次元の

ベクトルで表現されており、入力画像の畳み込みで得られた特徴量と fc4 層で結合されている。最終層である fc6 層では行動を表す確率分布を、fc7 層では信頼度をそれぞれ出力する。

5. 実験

本章では提案手法の有効性を確認するための実験についての記述、および実験結果についての考察を行う。一般物体データセットおよび細胞動画データセットを用いて、深度を考慮することによる有効性を確認するための実験を行った。

5.1 実験データ

5.1.1 一般物体画像のデータセット

一般物体動画画像を用いたネットワークの訓練のために、VoT2013 [17], VoT2014 [18], VoT2015 [19], ALOV300 [20] と呼ばれる 4 種類のデータセットを用いた。それぞれのデータセットには重複している動画画像が含まれているが、訓練時には重複を取り除き、合計 371 種類の一般物体動画画像を訓練に利用した。

5.1.2 細胞画像のデータセット

提案手法の学習と追跡精度の評価を行うため、皮膚の炎症を誘導したマウスの耳介を二光子励起顕微鏡によって観察/撮影した動画画像をデータセットとして用いた。これは励起光に 900 nm の波長を用い、446 nm, 525 nm, 575 nm 付近の蛍光を観察したものである。446 nm の波長に該当するのは、第二次高調波発生 (Second Harmonic Generation, SHG) で、皮膚ではコラーゲンの線維が特異的に可視化される。525 nm は緑色蛍光タンパク質 (EGFP) の波長に該当し、LysM-EGFP トランスジェニックマウスの顆粒球を可視化する。575 nm は特異的な対象はないが、あえていえば体毛等の自家蛍光が可視化される。これらのチャンネルのシグナルをそれぞれ、RGB 画像の青、緑、赤のチャンネルに割り振った疑似カラー画像として構成した。重要なチャンネルは緑であり、この緑色の細胞 (顆粒球) の遊走を追跡するのが本実験の目的である。

動画の解像度は XY 方向に 512×512 ピクセルであり、1 ピクセルは $0.5 \mu\text{m}$ に相当する。また、Z 方向には $3 \mu\text{m}$ 間隔で約 15 枚撮影したスタックである。1 本の動画は、1 分間隔の 20 フレームからなる。動画は 15 本あり、各動画ごとにそれぞれ 5 個程度の細胞についてアノテーションされており、訓練と評価に用いるシーケンスの数は 99 系列となっている。

5.2 実験設定

一般物体動画データセットを用いてネットワークを訓練した後、その重みを初期値として細胞動画データセットを用いて転移学習を行う。細胞動画の訓練および評価

表 1 実験に使用したパラメータ

Table 1 Parameters used for experiments.

パラメータ	値
教師あり学習エポック数 (一般物体)	30
強化学習エポック数 (一般物体)	30
教師あり学習エポック数 (細胞)	50
強化学習エポック数 (細胞)	10
矩形の移動倍率	0.03
学習率 (畳み込み層)	0.0001
学習率 (全結合層)	0.001
正クラスの矩形生成数 (Train)	150
正クラスの矩形生成数 (Test, フレーム $l = 1$)	200
正クラスの矩形生成数 (Test, フレーム $l > 1$)	30
負クラスの矩形生成数 (Train)	50
負クラスの矩形生成数 (Test, フレーム $l = 1$)	50
負クラスの矩形生成数 (Test, フレーム $l > 1$)	15
再検知と見なされる IoU の閾値	0.5
Fine-tuning が行われるフレーム間隔 (一般物体)	20
Fine-tuning が行われるフレーム間隔 (細胞)	5

のために、データセットを動画単位で 2 分割して交差検証を行った。表 1 に実験に使用したパラメータを示す。本手法で提案している追跡器は Python および深層学習フレームワークである PyTorch を用いて実装し、すべての実験は i7-6900K 3.20 GHz CPU, 128 GB RAM, Quadro GP100 GPU を搭載した計算機上で行われている。

5.3 評価手法

追跡器の性能を測るために、One-Pass Evaluation (OPE) [21] と呼ばれる手法を用いて Success rate と、Precision の 2 種類の指標に基づく評価を行う。Success rate は、あるシーケンスにおいて真値と予測結果との IoU が、与えられた閾値より高いフレームの割合を表す。Precision は、あるシーケンスにおいて真値と予測結果の中心間のユークリッド距離 (Center Location Error) が、与えられた閾値よりも小さいフレームの割合を表す。OPE では、評価に用いる全シーケンスについてこれらの値を、様々な閾値で計算しプロットする。また、本論文ではプロットから手法どうしを定量的に比較するために、Success rate においては Area Under Curve (AUC) を用いた。そして、追跡対象である顆粒球の大きさは約 $15 \mu\text{m}$ といわれているため、矩形が追跡対象を上手くとらえていることを示す基準として、Precision は閾値が $5 \mu\text{m}$ (10 ピクセル) のときの値を用いた。

5.4 実験結果と考察

図 6 は、実験で使用した細胞動画データセットから、3.2.1 項で説明した処理に従って生成された行動ラベル数の分布であり、“ADNet” [3] で一般物体追跡を行った際のラベルの分布と大きな差は見られなかった。このため、本

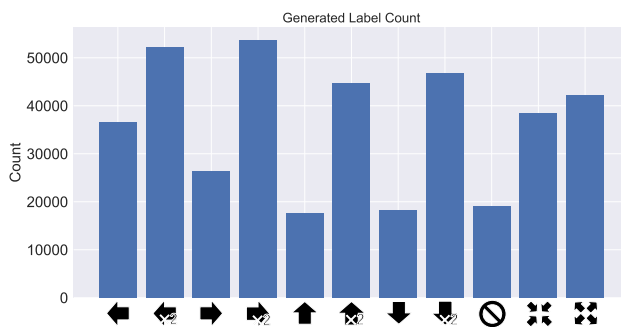


図 6 細胞動画データセットから生成された行動ラベルの分布
Fig. 6 Distribution of action labels generated from cell dataset.

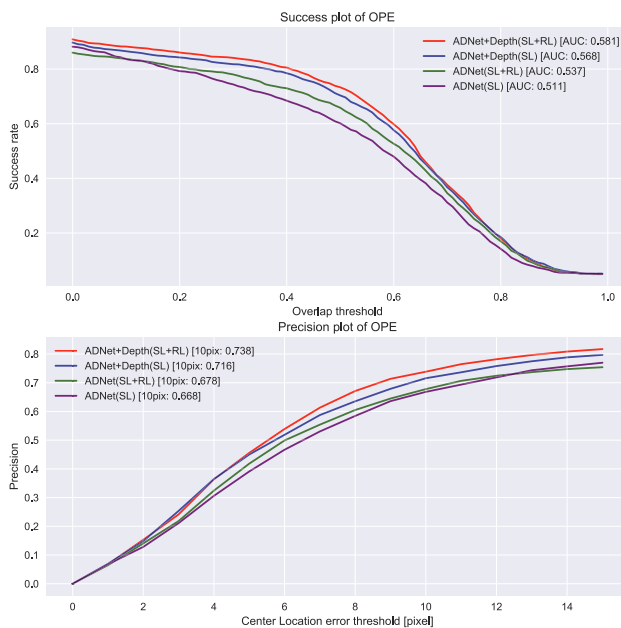


図 7 細胞動画データセットにおける Success plots と Precision による追跡手法の比較

Fig. 7 The success plots and precision of OPE on cell dataset.

実験の細胞動画による教師あり学習の際にも、特にラベル数の補正は行わず、3.2.2 項に従って訓練を行った。

図 7 は、細胞動画データセットを用いた 2-fold 交差検証による性能評価である。図 7 上のプロットは Success rate による評価を示しており、横軸は IoU の閾値を、縦軸は IoU がその閾値を超えたフレームの割合を表している。図 7 下のプロットは Precision による評価を示しており、横軸は 2 つの矩形の中心間のユークリッド距離の閾値を、縦軸は中心距離がその閾値未満であるフレームの割合を表している。凡例中の“ADNet+Depth”は本論文で提案する疑似深度を入力した ADNet を、“ADNet”は従来の ADNet を表している。また、“SL”は教師あり学習のみ、“SL+RL”は教師あり学習と強化学習を用いて訓練したネットワークによる追跡結果を表している。凡例中の各手法につけられた数値は、5.3 節で設定した評価値である。疑似深度を用いた“ADNet+Depth”は学習の方法にかかわらず、疑似深度を用いていない“ADNet”よりも Success

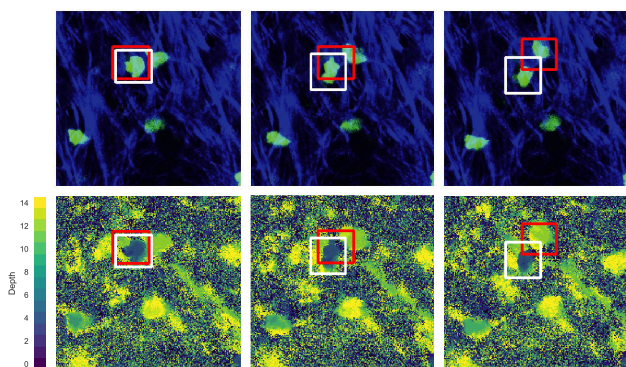


図 8 深度が異なる細胞の追跡

Fig. 8 Tracking on cells at different depth.

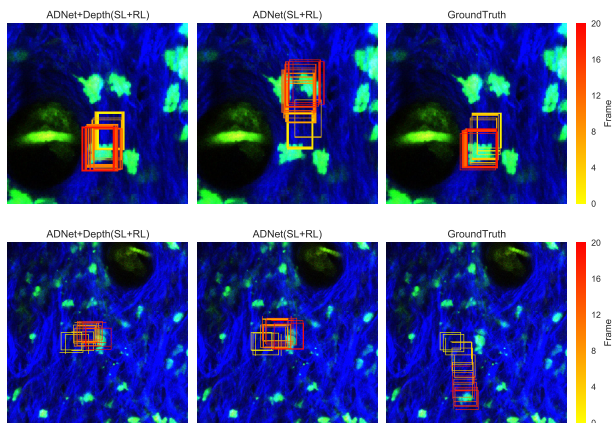


図 9 追跡結果の例

Fig. 9 Examples of tracking result and ground truth.

rate, Precision とともに性能が改善していることが分かる。

また、図 8 は、従来の“ADNet”で細胞の取り違えが発生した例である。赤色の矩形は従来の“ADNet”による追跡結果を表しており、白色の矩形は真値を表している。MIP 画像中の連続するフレームにおいて、細胞が非常に近接しており、従来の“ADNet”では対象を取り違えていた。しかし、疑似深度画像では、近接している細胞は疑似深度の値が異なっている。本手法では、深度情報を用いることで、このような 2 つの細胞が近接している状況下においても、それぞれの細胞を区別することが可能となっている。

図 9 は、追跡結果の例である。画像は最終フレームであり、1 フレーム目から最終フレームまでの矩形の追跡経路を画像上に出力している。疑似深度を用いることによって、図 9 上段のように周囲の細胞を誤認識せずに追跡が成功している。一方で、図 9 下段のように、追跡が失敗している例もあげられる。

追跡精度が低下する一因として、本手法では追跡を行う矩形のアスペクト比が固定されている点が考えられる。これは通常の矩形の移動予測においてはアスペクト比を変更するような行動を定義していないためである。細胞は一般物体とは異なり変形が頻繁に生じる。それに合わせて真値となる矩形も大きく変形することが多い。図 10 は、真値

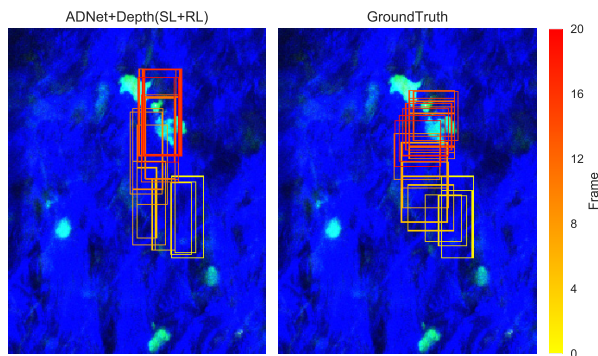


図 10 細胞追跡中におけるアスペクト比の変化

Fig. 10 Aspect ratio change during cell tracking.

となる矩形のアスペクト比の変化が大きいため、追跡精度が落ちてしまっている例である。真値では、縦に長い初期矩形が徐々に正方形に近づいているが、予測結果ではいずれの手法においても、アスペクト比がほとんど変化しておらず、2 つの細胞が矩形の中に含まれることがあり、細胞を取り違えてしまう可能性が高くなる。

6. おわりに

本論文では、蛍光動画像中の細胞追跡のために深度を考慮した深層学習手法を提案した。これは矩形の操作によって物体追跡を行う ADNet を拡張した方法である。細胞の追跡において、細胞動画像のみからなる訓練に十分なデータセットを準備することは困難であるが、一般物体追跡のタスクのために公開されているデータセットを用いて訓練したモデルから転移学習を行うことで、比較的少数の細胞動画像から細胞の追跡が可能になった。また、MIP によって 3 次元動画を 2 次元平面に投影する際に、本来であれば Z 軸方向の情報は失われてしまい、細胞が高密度に動画像内に写っているような場合の対応が難しいが、最大値の Z 座標を疑似的な深度としてニューラルネットワークに入力することで、Z 軸方向の情報を考慮した追跡を行い、精度が向上することを確認した。実験では一般物体動画像、細胞動画像のデータセットを用いた訓練をしたネットワークを用いて、細胞動画像データセットによる交差検証を行うことで、細胞追跡における疑似深度の有効性を確認した。

強化学習を取り入れた狙いとしては、過去の移動方法の履歴を入力に加えることから、細胞の移動傾向を学習することを期待したためであり、先行研究 [3] と同様に多少の精度向上が見られた。しかし、強化学習を用いても追跡が失敗する例を確認すると、矩形の内側に 2 つの細胞が含まれた時点において、取り違えが発生していることが多い。細胞が混雑した状況下では、対象の細胞の移動履歴のみでは追跡が困難であり、疑似深度の利用が効果的であると考えられる。

今後の課題としては、細胞の不規則な挙動に対してアスペクト比を追跡中に変形できるような操作を訓練すること

があげられる。加えて、細胞観察中には、細胞分裂や細胞の死による追跡対象の消滅といったような現象が確認できる場合があり、このような現象を合わせてとらえられるような行動を定義し、訓練することができれば、本研究の対象範囲がより広がると考えられる。

また、本研究ではコラーゲンや体毛といった背景のチャネルも加えた動画を対象に実験を行ったが、本手法において実質的に必要なのはEGFPと疑似深度のチャネルのみである。一方で、コラーゲンや体毛は動かない背景であると同時に細胞移動に対しての「道」や「障害物」になっており、そのような情報を有効に活用すれば、より正確に追跡を行える可能性がある。コラーゲンの繊維内で「道」を見ようとするときには、MIPによって3次元のトンネルがつぶれてしまうことが考えられるため、細胞の移動の自由度が周辺の物体によって強く制限されているような環境下では、追跡対象のチャネル以外の情報を上手く扱う方法も重要であると考えられる。

謝辞 本研究はJST CREST JPMJCR15G1の支援と、科研費18K19842, 18H05035の助成を受けたものです。

参考文献

- [1] Furuya, M., Kikuta, J., Fujimori, S., Seno, S., Maeda, H., Shirazaki, M., Uenaka, M., Mizuno, H., Iwamoto, Y., Morimoto, A., et al.: Direct cell-cell contact between mature osteoblasts and osteoclasts dynamically controls their functions in vivo, *Nature Communications*, Vol.9, No.1, p.300 (2018).
- [2] Kreisel, D., Nava, R.G., Li, W., Zinselmeyer, B.H., Wang, B., Lai, J., Pless, R., Gelman, A.E., Krupnick, A.S. and Miller, M.J.: In vivo two-photon imaging reveals monocyte-dependent neutrophil extravasation during pulmonary inflammation, *Proc. National Academy of Sciences*, Vol.107, No.42, pp.18073-18078 (2010).
- [3] Yun, S., Choi, J., Yoo, Y., Yun, K. and Young Choi, J.: Action-Decision Networks for Visual Tracking with Deep Reinforcement Learning, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
- [4] Kumagai, S. and Hotta, K.: Particle detection in intracellular images and radius estimation by circle fitting, *IEEJ Trans. Electrical and Electronic Engineering*, Vol.10, No.2, pp.181-185 (2015).
- [5] McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al.: CellProfiler 3.0: Next-generation image processing for biology, *PLoS Biology*, Vol.16, No.7, p.e2005970 (2018).
- [6] Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäkel, Z., Seiwald, K., et al.: U-Net: Deep learning for cell counting, detection, and morphometry, *Nature Methods*, Vol.16, No.1, p.67 (2019).
- [7] Bise, R., Yin, Z. and Kanade, T.: Reliable cell tracking by global data association, *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp.1004-1010, IEEE (2011).
- [8] Yang, X., Li, H. and Zhou, X.: Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy, *IEEE Trans. Circuits and Systems I: Regular Papers*, Vol.53, No.11, pp.2405-2414 (2006).
- [9] Hirose, O., Kawaguchi, S., Tokunaga, T., Toyoshima, Y., Teramoto, T., Kuge, S., Ishihara, T., Iino, Y. and Yoshida, R.: SPF-CellTracker: Tracking multiple cells with strongly-correlated moves using a spatial particle filter, *IEEE/ACM Trans. Computational Biology and Bioinformatics (TCBB)*, Vol.15, No.6, pp.1822-1831 (2018).
- [10] Kristan, M., Matas, J., Leonardis, A., Vojir, T., Pflugfelder, R., Fernandez, G., Nebehay, G., Porikli, F. and Čehovin, L.: A Novel Performance Evaluation Methodology for Single-Target Trackers, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.38, No.11, pp.2137-2155 (2016).
- [11] Ning, G., Zhang, Z., Huang, C., He, Z., Ren, X. and Wang, H.: Spatially Supervised Recurrent Convolutional Neural Networks for Visual Object Tracking, arXiv preprint arXiv:1607.05781 (2016).
- [12] Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al.: Mastering the game of Go with deep neural networks and tree search, *nature*, Vol.529, No.7587, p.484 (2016).
- [13] Williams, R.J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning, *Machine learning*, Vol.8, No.3-4, pp.229-256 (1992).
- [14] Nam, H. and Han, B.: Learning Multi-Domain Convolutional Neural Networks for Visual Tracking, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4293-4302 (2016).
- [15] Chatfield, K., Simonyan, K., Vedaldi, A. and Zisserman, A.: Return of the devil in the details: Delving deep into convolutional nets, arXiv preprint arXiv:1405.3531 (2014).
- [16] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. and Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge, *International Journal of Computer Vision (IJCV)*, Vol.115, No.3, pp.211-252 (2015).
- [17] Kristan, M., Pflugfelder, R., Leonardis, A., Matas, J., Porikli, F., Čehovin, L., Nebehay, G., Gustavo, F. and Vojir, T.: The Visual Object Tracking VOT2013 challenge results, *Workshop on the VOT2013 Visual Object Tracking Challenge*, pp.98-111 (2013).
- [18] Kristan, M., Pflugfelder, R., Leonardis, A., Matas, J., Čehovin, L., Nebehay, G., Vojir, T. and Fernandez, G.: The Visual Object Tracking VOT2014 challenge results, *Workshop on the VOT2014 Visual Object Tracking Challenge*, pp.191-217 (2014).
- [19] Kristan, M., Matas, J., Leonardis, A., Felsberg, M., Čehovin, L., Fernandez, G., Vojir, T., Hager, G., Nebehay, G. and Pflugfelder, R.: The visual object tracking vot2015 challenge results, *Proc. IEEE International Conference on Computer Vision Workshops*, pp.1-23 (2015).
- [20] Smeulders, A.W., Chu, D.M., Cucchiara, R., Calderara, S., Dehghan, A. and Shah, M.: Visual tracking: An experimental survey, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.36, No.7, pp.1442-1468 (2014).
- [21] Wu, Y., Lim, J. and Yang, M.-H.: Online object track-

ing: A benchmark, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.2411-2418 (2013).



嶋田 彩人

2017年大阪大学基礎工学部卒業。2019年大阪大学大学院情報科学研究科修了。2019年よりソニー株式会社に勤務。



瀬尾 茂人 (正会員)

2006年大阪大学大学院情報科学研究科修了。博士(情報科学)。同年同研究科助手。2007年助教。2017年より大阪大学大学院情報科学研究科准教授。日本バイオインフォマティクス学会各会員。



繁田 浩功 (正会員)

2016年大阪大学大学院情報科学研究科修了。博士(情報科学)。2015年10月大阪大学サイバーメディアセンター特任研究員。2016年大阪大学大学院情報科学研究科特任研究員。2017年より大阪大学大学院情報科学研究科特任助教。バーチャルリアリティ学会, ACM 各会員。



間下 以大 (正会員)

大阪大学大学院基礎工学研究科博士後期課程修了。現在, 同大学サイバーメディアセンター准教授。2012年オーストリア・グラーツ工科大学客員研究員。コンピュータビジョン, 機械学習, 拡張現実に関する研究に従事。博士(工学)。



内田 穂

助教。

2007年横浜市立大学大学院医学研究科修了。博士(医学)。2007年横浜市立大学大学院医学研究科特任助教。2008年米国立衛生研究所心臓・肺・血管研究所 Visiting Fellow。2016年より大阪大学大学院医学系研究科特任



石井 優

教授。2013年より大阪大学大学院医学系研究科/生命機能研究科教授。

1998年大阪大学医学部医学科卒業。2000年大阪大学大学院医学系研究科助手(2005年医学博士)。2006年米国立衛生学研究所客員研究員。2009年大阪大学免疫学フロンティア研究センター特任准教授。2011年同特任教



松田 秀雄 (正会員)

1987年神戸大学大学院自然科学研究科修了(学術博士)。同年同大学工学部助手。1994年大阪大学基礎工学部助教。2002年より大阪大学大学院情報科学研究科教授。バイオインフォマティクスの研究に従事。日本バイオインフォマティクス学会, ISCB, IEEE CS, ACM 各会員。