

機械学習を用いた密度マップ推定による大規模魚群中の個体検出

Small Fish Detection in Large Fish School by Density Map Estimation Using Machine Learning

村田 一星† 波部 斉† 阿部 孝司† 井口 信和†
Issei Murata Hitoshi Habe Koji Abe Nobukazu Iguchi

1. はじめに

養殖場や水族館において、生育状況や飼育環境の把握のために魚の尾数を計測し、適切な飼育環境を管理することが重要である。特に本稿で検出対象とするクロマグロの養殖場では頻繁に尾数を計測している。その理由は、マグロが瞬発的に移動し、壁に衝突する原因とされる周囲の突発的な音や光を解明することや、餌の量による生存率を確認しているためである。現在、尾数計測は人手によって柄杓とバケツを用いて行われている。この方法では、魚にストレスや外傷を与え死滅の原因となる可能性が高い。さらに、1匹ずつ人手で計測しているため人手による負担も大きい。そこで、水槽をカメラで撮影し、その映像からマグロの尾数を計測することで魚の管理ができれば、魚に対してストレスや外傷を負わせることなく、人手による作業量も軽減することができ、より効率的に業務を行うことができると考えられる。また、自動的に大量の尾数を管理することができるようになれば、リアルタイムで状況を把握することができる。そうすることで、尾数の急激な変化も検知でき、魚群の異常行動や死亡原因の解明にも役立つと期待される。本稿では大規模魚群中の個体検出として、クロマグロの稚魚を対象として実験を行い、評価を行う。

2. 関連研究

画像中から人や車両、細胞などを検出・カウントする手法は長年にわたって研究されてきた[1][2][3]。映像中の魚の検出・追跡や分類においても様々な手法が提案されている[4][5][6][7]。その多くは比較的大きく映った数匹から数十匹の魚を対象とするものである。本研究と特に関連するものとして、Maらは、小さく写った数十匹の様々な生物を対象とし、画像特徴量に基づく密度推定と整数計画法を組み合わせた高精度な個体検出手法を提案している[8]。しかし、養殖場などの水槽には数千匹の微小な魚が遊泳しており、それらを検出する手法が必要となる。過去に我々がCNNを用いた密度推定による大規模魚群の個体検出手法を提案し、クロマグロ稚魚を高精度に検出している[11]。しかし、実際の養殖現場で要求される精度に満たすことができておらず、また学習時に大量のラベル付きデータを要するなどの問題がある。本稿では、より高精度な個体検出を実現する技術を提案する。

3. 提案手法

本研究では、実際の養殖現場で利用できる高精度な魚の検出技術を目指とする。CNN(畳み込みニューラルネットワーク)[12]を用いることで人間が思いつかないような特徴量を取得することができ、過去に我々が提案したCNNのみによる手法で高精度を示している。しかし、CNNのみによる手法では大量のラベル付きデータによる学習が必要であり、現実的ではない。そこで、予測に必要な境界

線付近の一部のデータのみによって高精度な識別を可能とするSVM(サポートベクトルマシン)をベースにした回帰分析手法であるSVR(サポートベクトル回帰)を採用した。そうすることで、少量のラベル付きデータによる学習で高精度を期待できる。また、CNNにより高い精度で検出が可能であったことを考慮し、特徴抽出器にはCNNを使用する。つまり、CNNによる特徴抽出とSVRによる高精度な個体検出を実現する技術を提案する。まず、CNNとSVRについて述べた後で、本手法で検出に使用する密度マップと、CNNとSVRによる密度マップ推定について述べる。

3.1 CNN(畳み込みニューラルネットワーク)

CNNとはコンピュータビジョンの分野で主に用いられているニューラルネットワークであり、畳み込み層(Convolution Layer)と、プーリング層(Pooling Layer)を多層に重ねて構成される。

畳み込み層では、入力画像にフィルタをスライドさせ画像の濃淡パターンを抽出し、特徴マップを作成する。画像全体から対象物体の特徴を抽出することができたため、位置の変化に頑健である。例えば、画像の左側で特徴を抽出して学習すれば、検出する際に対象物体が画像の右側にある場合でも検出することができ、汎用性がある。

プーリング層では、畳み込み層で作成した特徴マップ上にウィンドウをスライドさせ重要な値だけを残す処理を行う。そうすることで、情報を圧縮して計算量を少なくすることや、入力画像のノイズや変形を吸収することができる。処理手法として、スライドさせるウィンドウ内の特徴量の最大値を残す最大値プーリングや、平均値を残す平均値プーリングと呼ばれる手法が用いられている。本稿では最大値プーリングを使用する。

3.2 SVR(サポートベクトル回帰)

SVRは、SVM[13][14]をベースとした回帰分析手法である。SVMとは、教師あり学習に基づいた機械学習手法であり、クラス分類や回帰に適用でき、現在知られている手法の中でも認識能力が優れた識別器とされている。また、SVMでは識別境界を引く際に、他クラス間で最も近い距離にあるデータ同士(サポートベクトル)を基準として、それらの距離間が最大になるような位置に境界線を設定するマージン最大化が導入されているため汎化性能が高く、SVRにも高い汎化性能が期待される。また、SVRはカーネル関数を使用しているため、高次元空間にデータを変換することができ、データの分離が容易である。カーネル関数には、RBFカーネル、線形、多項式などがあり、今回は非線形で使用できるRBFカーネルを採用する。RBFカーネルは以下のように定義される。ここで、 $\|x - x'\|^2$ はユークリッド空間上の距離の2乗である。 $\gamma = \frac{1}{2\sigma^2}$ であり、訓練データが影響を及ぼす範囲をパラメータとして与えることができる。

$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

† 近畿大学, Kindai University

SVR による回帰式は以下のとおりである． a_i , \hat{a}_i , b は学習の結果として得られるパラメータである．

$$y = f(x) = \sum_{k=1}^n (a_i - \hat{a}_i) K(x, x_i) + b$$

また，パラメータには誤りに対する厳しさを表す C を与えることができる． C が小さいほど誤りに対して甘くなり，大きくなるほど厳しくなるが過学習の原因となる．

3.3 密度マップ

密度マップとは，入力画像のピクセル毎に検出対象個体が存在する確からしさである確信度を表す密度である．したがって，密度マップは入力画像と同じサイズとなる．密度マップの例を図1に示す．密度マップ中で個体が存在しない位置は確信度が最小値である．そのため，個体位置に付与するポジティブデータのラベル付けのみで，ネガティブデータである背景を容易に学習させることができる．入力画像の各ピクセルと対応した正解とする密度マップは [9][10] と同様に，物体に付与する位置情報を用いて以下の式で表されるガウシアンフィルタに従って生成した．

$$F(x) = \sum_{p \in P_{gt}} P_{\sigma}(x - p)$$

x は画像中の座標， P_{gt} は個体の正解位置の集合， P_{σ} は幅 σ のガウシアンカーネルである． x と P_{gt} の各正解ピクセルとの距離 $(x-p)$ によって重みを変えながらぼかし処理を行い，密度マップを作成する．

3.4 CNN による特徴抽出と SVR による密度マップ推定

図2にCNNによる特徴抽出とSVRによる検出手法を示す．入力画像の各ピクセルに対応する密度マップの値を学習する際には，各ピクセルを中心とした縦横サイズ30の局所画像を用いて特徴を抽出する．その後，全結合(Fully Connected)層から特徴量を取得し，局所画像に対応する密度マップの出力値をSVRにより学習する．特徴抽出のためのCNN構造は図2のとおりである．層の構成については，高精度な検出かつ汎用性の向上や検出時間の短縮を可能にするため，少ない層での構成を目指した．

検出する際には，検出用の魚群画像に対して，学習時と同様に局所画像を用いて各ピクセルに対応する密度を学習されたSVRにより推定し，密度マップを作成する．作成された密度マップで密度が隣接して生成された部分の中からNMS(Non-Maximum Suppression)を適用し，確信度が最も高い箇所以外を最小値に抑え1点のみを抽出し，稚魚の座標を検出する．

4. 実験

4.1 データセット

近畿大学水産研究所大島実験場で取得したデータセット(Bluefin Tuna Dataset, 以下BTD)を用いる．BTDは図3に示すように円形の水槽上部に設置したカメラにより撮影したデータである．水槽の直径は6m，水深は1mである．水槽内部には体長数cm程の微小なクロマグロの稚魚が約7000匹遊泳している．撮影した映像の中から3フレームを

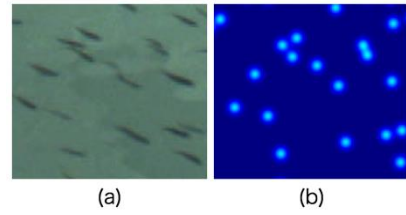


図1 密度マップ例

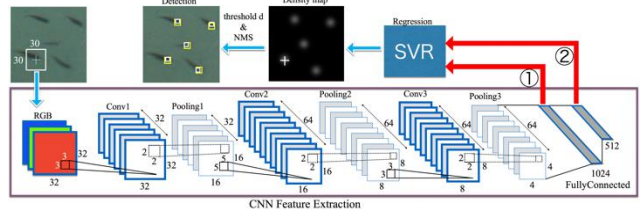


図2 提案手法の流れ

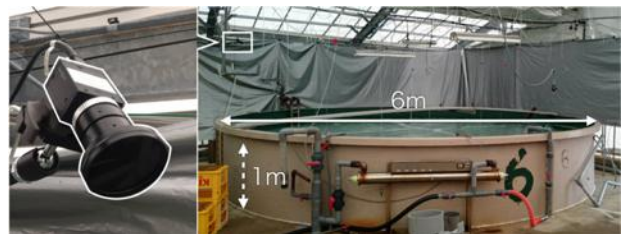


図3 撮影環境

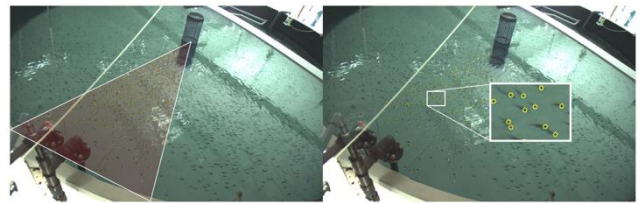


図4 カメラで撮影したクロマグロの魚群
(a) 実験で対象とする画像領域
(b) 目視で特定した個体位置

取り出し，1フレームを学習用の画像，2フレームを検出用の画像として使用した．また，図4(a)に示す領域内で泳ぐ稚魚を検出対象とした．図4(b)に示すように，頭部を目視で特定したものを正解データとして使用した．

4.2 実験手順

検出を行う前の条件として，魚群画像に稚魚の座標位置を与えるアノテーションと呼ばれる作業を行い，ラベル付きデータを取得しておく必要がある(図4(b))．本手法では1フレームに写る全ての魚に対してアノテーションを行い，取得された座標データから魚が存在する確信度を表す密度マップを作成し，学習用に使用する．また，SVRでの学習時に定めるパラメータはグリッドサーチを用いて最適な値を求めた．

本実験では，先行研究であるCNNを特徴抽出器と識別器として用いた検出手法による精度と，CNNの様々な層(図2①②)から特徴量を抽出しSVRによる識別を組み合わせた手法との精度を比較する．図2①はCNNによる特徴抽出が終了した段階で取得する特徴量である．図2②はCNNによる特徴抽出と識別途中の段階で取得する特徴量で

ある。これらにより、CNN や SVR による識別の性能を評価する。学習用の画像として 1 枚選択しその画像内から、稚魚とそれに対応する密度マップ値の組を 9828 サンプル、背景とそれに対応する密度マップ値の組を 9828 サンプル、計 19656 サンプルを取得して学習を行う。また、実験評価用の画像として 2 枚選択する。評価値には再現率、適合率、F 値を用いる。

4.3 実験結果

BTD に対して提案手法を適用した結果を表 1 に示す。2 枚の実験評価用の画像に対して、それぞれの手法を適用した結果の再現率、適合率、F 値を示し、F 値のみ 2 枚の平均も示している。また、検出結果の一例を図 5、図 6、図 7 に示す。検出結果の図（図 5、図 6、図 7 の右側）において、白枠は正解位置、黄枠は検出位置である。実験結果より、全ての手法において F 値は 80% を超える高い数値であることがわかる。CNN のみを用いた手法では適合率は高いが、再現率は比較的低い。図 5 の右上拡大図から、水面の反射などの影響による誤検出が多数見られる。次に、図 2①に示す CNN による特徴抽出後の段階の層から得られた特徴量から SVR による識別を行った結果では、F 値が 88% という高精度な検出結果が得られた。図 6 の水面の反射がある部分を確認しても誤検出箇所が少なく、概ね正しく検出されていることが分かる。最後に、図 2②に示す CNN による識別途中の段階の層から得られた特徴量から SVR による識別を行った結果では比較的高精度な結果が得られたが、特徴抽出後の段階の層（図 2①）から特徴量を抽出し、SVR による識別を行った結果よりも精度が低い結果となった。

表 1 検出結果

	検出画像	再現率 (%)	適合率 (%)	F 値 (%)	F 値の平均 (%)
CNN	1 枚目	76.7	87.8	81.9	80.2
	2 枚目	73.0	84.9	78.5	
CNN+SVR (図 2①)	1 枚目	89.4	87.8	88.6	88.0
	2 枚目	87.9	86.6	87.3	
CNN+SVR (図 2②)	1 枚目	85.8	89.9	87.8	84.9
	2 枚目	79.3	84.9	82.0	

4.4 考察

本研究は高精度な大規模魚群中の微小個体検出を目標として、CNN による特徴抽出と SVR による識別を組み合わせる手法を提案し、実験を行った。実験では、同数のラベル付きデータによる条件下の学習において、CNN のみによる手法と比べ、CNN と SVR を組み合わせる手法に高い精度が得られている。このことから、提案手法はより高精度に検出が可能であり、実際の養殖場での業務効率の向上に期待できる。また、CNN による特徴抽出後の段階で特徴量を取得し（図 2①）、検出を行った場合に、CNN の識別途中から特徴量として取得した場合（図 2②）よりも高精度な検出結果が得られた。以上のことから、識別部には CNN よりも SVR を用いることで優れた識別を可能とすることが分かる。

また、実際の養殖現場では、検出尾数と実際の尾数との誤差が 10% 以内であれば運用できるが満たすことができていない。これは、学習時に与えるデータ数不足や水面による反射、さらに水深による魚影の不鮮明さが影響していると考えられる。このことは、検出結果の画像から水面が比較的綺麗な箇所ではどの手法においても誤検出が少ないことから言える。

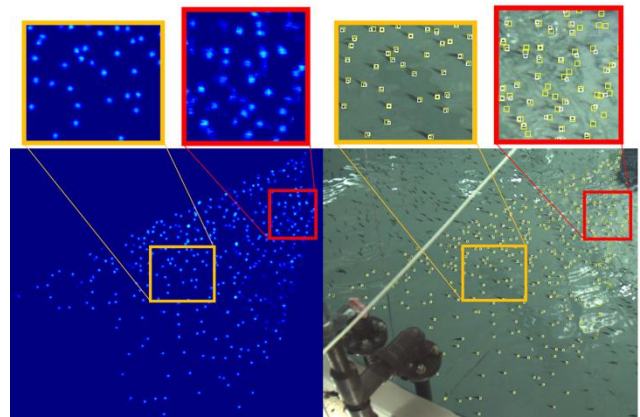


図 5 CNN 検出結果の一例

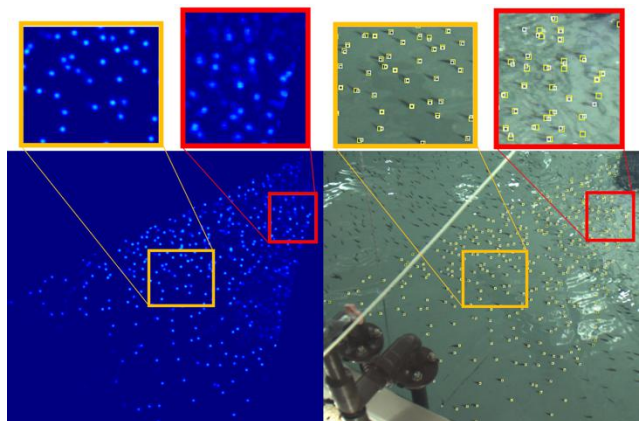


図 6 CNN+SVR (図 2①) 検出結果の一例

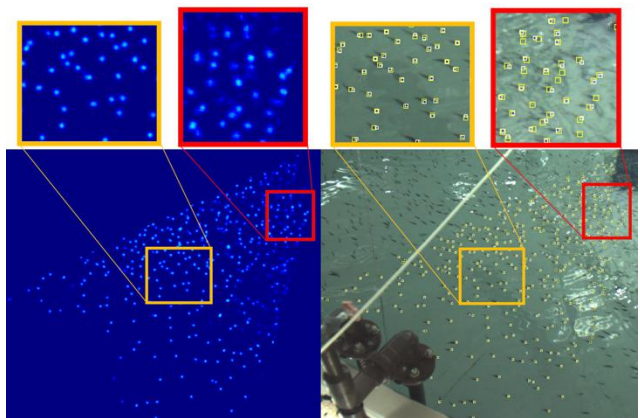


図 7 CNN+SVR (図 2②) 検出結果の一例

5. まとめ

CNN を用いた特徴抽出と SVR を用いた識別を組み合わせた大規模魚群中の高精度な個体検出手法を提案し、実際の養殖場で撮影した映像を用いて実験を行った。その結果、同数のラベル付きデータで学習した際に、CNN のみで特徴抽出と識別を行った場合よりも提案手法に再現率、適合率、F 値が上回る結果を得ることができ、より実際の養殖現場での業務効率の向上に繋がる結果となった。今後は学習データ数不足を補いより高い業務効率を目指すために、自動でデータを回収できるようにすることが課題であると考えている。また、学習時や検出時に要する時間を短縮化している。

など、実用化に向けた取り組みも課題とする。

謝辞

本研究を行う上で貴重なご意見を頂戴し、映像取得にご協力いただいた近畿大学水産研究所の升間主計所長、近畿大学水産養殖種苗センターの岡田貴彦センター長、ならびに奄美実験場の皆様に御礼を申し上げる。本研究の一部は、JSPS 科研費 JP17H05981, JP17K00256 の助成を受けて行った。

参考文献

- [1] Arteta, C., Lempitsky, V., Noble, J. A. and Zisserman, A., “Learning to Detect Partially Overlapping Instances”, CVPR, pp. 3230-3237, 2013.
- [2] Lempitsky, V. and Zisserman, A., “Learning To Count Objects in Images”, NIPS, 2010.
- [3] Ryan, D., Denman, S., Sridharan, S. and Fookes, C., “An Evaluation of Crowd Counting Methods, Features and Regression Models”, CVIU, Vol. 130, pp. 1-17, 2015.
- [4] Boom, B. J., Huang, P. X., Beyan, C., Spampinato, C. et al., “Long-Term Underwater Camera Surveillance for Monitoring and Analysis of Fish Populations”, VAIB, 2012.
- [5] Fan, L. and Liu, Y., “Automate Fry Counting using Computer Vision and Multi-Class Least Squares Support Vector Machine”, Aquaculture, Vol. 380, pp. 91-98, 2013.
- [6] French, G., Fisher, M., Mackiewicz, M. and Needle, C., “Convolutional Neural Networks for Counting Fish in Fisheries Surveillance Video”, BMVA, 2015.
- [7] Spampinato, C., Chen-Burger, Y.-H., Nadarajan, G. and Fisher, R. B., “Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos”, VISAPP (2), Vol. 2008, pp. 514- 519, 2008.
- [8] Ma, Z., Yu, L. and Chan, A. B., “Small Instance Detection by Integer Programming on Object Density Maps”, CVPR, 2017.
- [9] Lempitsky, V. and Zisserman, A., “Learning To Count Objects in Images,” NIPS, 2010.
- [10] Ma, Z., Yu, L. and Chan, A. B., “Small Instance Detection by Integer Programming on Object Density Maps,” CVPR, 2017.
- [11] 村田一星, 佐藤僚太, 波部斉, 阿部孝司, 井口信和 “密度マップ推定による大規模魚群中の個体検出,” 第 24 回画像センシングシンポジウム (SSII2018) , 2018.
- [12] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., “Gradient-Based Learning Applied to Document Recognition,” IEEE, Vol. 86, No. 11, pp. 2278-2324, 1998.
- [13] V.Vapnik and A.Lerner, ”Pattern recognition using generalized portrait method,” Automation and Remote Control, 24, 1963.
- [14] Bernhard E. Boser, Isabelle M. Guyon, Vladimir N. Vapnik, ”A training algorithm for optimal margin classifiers,” Proceedings of the fifth annual workshop on Computational learning theory –COLT’92, p.144, 1992.