The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

CNN を用いたサケの雌雄判別の試み

Trial to determine the sex of salmon using CNN

阿部 周司 *1　　石畑 宏明 *2
Shuji ABE　　Hiroaki ISHIHATA

*1東京工科大学応用生物学部
School of Bioscience and Biotechnology, Tokyo University of Technology

*2東京工科大学コンピュータサイエンス学部
School of Computer Science, Tokyo University of Technology

We tried to deteaminate the sex of salmon using a convolutional neural network(CNN). We collected labeled(male/female) salmon image data for two years, then trained VGG 16 type network. High accuracy was obtained with combination of learning data and test data of the same year. When the learning data and the test data are different year, it is almost impossible to recognize it. By using both data for learning, high recognition accuracy was obtained. It is important to obtain various variety of data not only for fish but also for the background. We also investigated observation point of CNN using Gradcam.

1. はじめに

ニューラルネットワークによる画像のクラス分類は，Convolutional Neural Network(以下 CNN) とそれに関する新技術の登場以後大きく進歩した．1000 カテゴリへの画像分類では，ImageNet，VGG，GoogleNet など新しいモデルが提案され，認識率は向上しその精度は人間を超えたとの報告もある．

本論文では，CNN を用いてサケの雌雄判別を実施した結果を示す．筆者らは，2017 年度に同様の研究を実施・報告した [阿部 18] が，今回はその続報となる．サケの画像データから，その個体が雌雄どちらのクラスに属するかを判別する．2 クラス分類なので容易に思えるが，サケの雄雌の外観の特徴は微妙であり，この違いが見分けられるかが課題である．これまで，2 クラスの分類は，人の性別 [Verma 14, Levi 15] など多数行われているが，その多くが，人の顔のようにバリエーションが多いものを分類するところに特徴がある．サケの雌雄判別では，見た目の差が微妙な違いを分類するという点で異なる困難さが予想される．

2 節ではサケの雌雄判別方法について，3 節ではラベル付きデータの収集方法，データアーギュメンテーションについて，4 節では，ニューラルネットの構成や学習方法，実験とその結果について述べる．

2. サケの雌雄判別

漁業者が捕獲したサケは雄と雌で価格が異なる．このため，漁業者は捕獲したサケを雌雄分別後，加工工場などに出荷する事が多い．現状，分別作業は，熟練者の人手によって行われている．

漁業者から聞き取りをしたところ，サケの雌雄判別は，図 1 に示すように，その頭 (A) 部，脂鰭 (B)，尾鰭 (C) の形状で判別している．実際には，これ以外に全体の形や感触も情報として利用しているように感じた．また，サケはその捕獲場所や捕獲時期により雄雌の形状の違いが変化してくる．河川を遡及する前は未成熟で雌雄差は少ないが，成熟するに従い差が大きくなる．雌雄判別においては，このような水揚げ時期に伴う対象の変化にも対応できる必要がある．

阿部 周司　東京工科大学応用生物学部　〒 192–0982 東京都八王子市片倉町 1404–1, abeshj@stf.teu.ac.jp

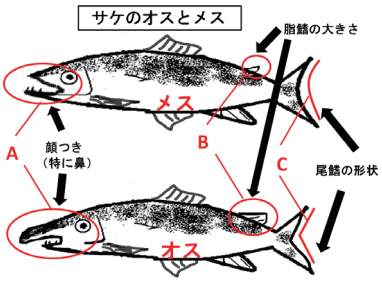

図 1: サケの雌雄の判別ポイント

3. データ収集

3.1 画像データ収集方法

ニューラルネットワークを使うため大量のラベル付き画像が必要となる．水産加工場に協力を依頼し，加工工場内に図 2 に示すような撮影機材を設置し，加工するサケを一時借用して撮影した．

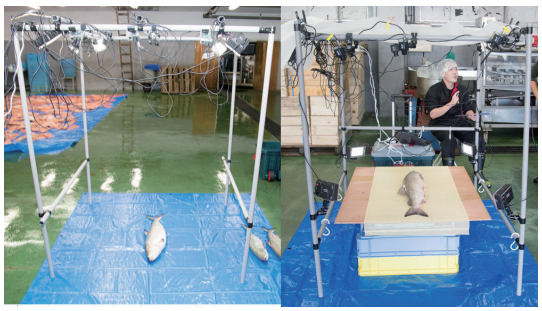

図 2: 画像データ取得システム (左:2017 年 右:2018 年)

2017 年は，床にブルーシートを敷き水槽から取り出したサケを配置して，12 個のカメラで 12 方向から同時に撮影した．

サケを表裏反転し再度撮影するという手順で 1 匹あたり 2 回，24 枚の静止画を取得した．機材のコスト削減のために，低価格な Web カメラを使用した．

2018 年度は前年の経験を踏まえ，魚体とカメラの距離を近づけられるようにカメラをより広角のものに代えた．魚体をマットを敷いた台の上に配置できるようになり，魚体の取り出し・配置の労力を大幅に削減した．一度に撮影する方向を 9 方向とし，1 匹あたり 18 枚の静止画を取得した．

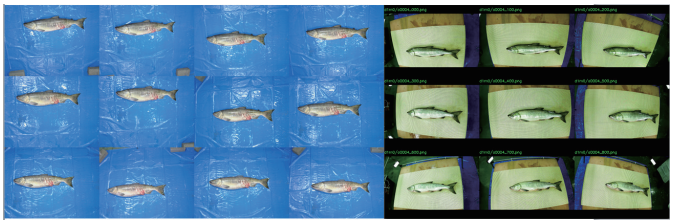

図 3: 取得した画像の例 (左:2017 年 右:2018 年)

2017 年度は，3 日間でオス 378 匹，メス 73 匹の撮影を行い，オス画像 9072 枚メス画像 1752 枚を取得した．メスは撮影当日が荒天のため魚の入荷がなくサンプル数が少なくなった．2018 年度は，2 日間でオス 584 匹，メス 411 匹を撮影し，オス画像 10512 枚メス画像 7398 枚を取得した．背景は水に濡れて強い反射を生むブルーシートではなく，黄色いマット状の布を用いて照明による強い反射を防いだ．照明の方向も上からではなく横から当てるようにした．

最終的に，表 1 に示すようなデータを取得した．なお，雌雄のラベル付けを確実に行うために，オスとメスを混在させることなく別々に撮影した．撮影を繰り返すと，ブルーシートやマット上にに水たまりの跡や汚れが付く．時々清掃を行ったが背景にそのようなノイズが生じることは避けられなかった．

後で述べるように，ニューラルネットワークには，320 × 240 に縮小した画像を入力する．2017 年と 2018 年のデータでは画像のアスペクト比が異なるので，2018 年のデータは上下に黒の余白を設けて，2017 年のものに合わせた．

表 1: 取得した画像データ

取得日時	2017/10/10-12		2018/9/23-24	
使用カメラ	Logicool c270		ELP 130W01MT	
解像度	640 × 480		1280 × 720	
性別	オス	メス	オス	メス
個体数	378	73	584	411
総データ数	9072	1752	10512	7398
DS0	504	288	1008	1008
DS1	1396	432	900	900
DS2	920	304	927	909
DS3	984	288	918	900

3.2 データセット

取得したデータを，4 通りの分割方法で学習データとテストデータに分割し，データセット DS0∼3 とした．DS0 は，2017 年データについては [阿部 18] で用いた分割である．2018 年データは，雄雌両方共撮影順で 1008 枚 (56 匹分) をテストデータとし，残りを学習データとした．残りの DS1∼3 は，撮影中の途中の部分を適当な数だけ取り出しテストデータとし，残りを学習データにした．どの場合も，同一魚体がテストデータと学習データに現れることはないように分割した．

2017 年度の実験結果から，背景の影響を軽減するためには魚体部分のみを切り出す必要があることが判明していたのでそのためのデータを作成した．2017 年のデータは，8562 枚について切り出し処理を人手でおこなった．2018 年のデータは，877 枚について切り出し処理を行った．残りのデータについては，人手による領域分割結果を教師データとして別のニューラルネットワークをトレーニングし，それを用いて領域データを作成した．目視による検査でほぼ正確に切り出しが行えていることを確認した．

3.3 データアーギュメンテーション

学習時には，表 2 に示すデータの水増し (data argumentation 以下 DA) を行った．表中の確率の列に示した確率で，種々の画像の変形を行った．項目 1 は背景の影響を軽減するために，魚体以外部分を灰色一色もしくは乱数データで埋めるものである．項目 2 以降で，平行移動，ランダムに変形, 上下反転, 回転，明度変更，ノイズ付加をこの順序で行った．DA の処理は，学習時にその都度実行した．従って，同一の画像データに対しても異なる学習データとしてネットワークに与えられる．

表 2: データアーギュメンテーション (DA) の内容

項目	確率 [%]	実施内容
1	80	魚体を含む長方形を切り出し，背景を 50%の確率でグレー・一様乱数とする．
2	50	画像全体を縦横サイズの 10%以下のランダムな距離分平行移動する．
3	70	画像の縦横サイズの 10%以内の範囲で四隅のランダムな位置を設定し，その 4 点を結ぶ四角形を長方形に射影変換によって変形する．
4	50	上下・左右それぞれ 50%の確率で反転する．
5	80	±10 度の範囲で魚体を中心にランダムに回転する．
5	20	画像の明度を ±50%変化させる．
7	10	平均 128，標準偏差 12.8 のガウシアンノイズを付加する．

4. 実験

4.1 ニューラルネットワーク

雌雄判別を行うニューラルネットワークとして，VGG16 を改変したものを使用した．ネットワーク構造の変更点は，1) 2 次元 convolution 後に batch normalization 処理を挿入したこと，2) オリジナルの fc6, 7 の全結合ネットワークのチャネル数を 1024 に，出力を 2 クラス分類に合わせて 2 出力としたことである．入力となる画像データは，320×240 画素のサイズで，0∼255 の RGB データを，それぞれ ±1.0 の範囲に変換したものとした．オリジナルの VGG16 の様に入力データから平均値を引く処理は行わなかった．

ニューラルネットワークの処理フレームワークとして chainer を使用した．オプティマイザとして Adam を $\alpha = 10^{-5}$ で使用し，30 エポックの学習を行った．途中 16 エポック目から $\alpha = 10^{-6}$ にした．その他のハイパーパラメータは全て chainer のデフォルトを使用した．学習と評価には，GTX-1080 搭載のワークステーションを使用した．

4.2 学習

表 1 にある 4 種類のテストデータ分割方法それぞれについて，ニューラルネットワークの学習を行った．メスデータについては，オスデータと同数になるように，データをコピーして追加した．以後，2017 年のデータのみを使用して学習したものを Y7 モデル，2018 年のデータのみを使用して学習したものを Y8 モデル，両方を使用して学習したものを Y0 モデルと呼ぶ．

DA を行ったテストデータ (DA あり) と DA を行わない撮影したままのオリジナルテストデータ (DA なし) について，学習後のネットワークを用いて判別を行い，精度と再現率から F 値を求めた．予測結果が全て片方になり，精度や再現率が算出できない場合は，F 値は 0 とした．その結果を表 3 に示す．表中の F 値はオスとメスの F 値を別々に算出し，その平均値をとっている．

Y7 モデルでは，DS0, DS3 のデータ分割方法で，低い F 値となった．Y8 モデルでは，いずれの分割方法でも 98%以上の高い F 値が得られた．先にも述べたように，Y7 で学習に使用したデータはメスの数が圧倒的に少なく，データの分割方法により大きな差異が出たと考えられる．Y0 モデルも，DS3 以外の分割方法で高い F 値が得られている．データのバラェティが増えたためと考えられる．DS3 のケースでは，2017 年のデータの分割方法により全体の F 値が低下していると考えられる．また，DA を行ったテストデータの場合よりも，DA しないデータの方が高い F 値となっている．

表 3: 各 DS の F 値 [%]

	DA ありデータ			DA なしデータ		
DS	Y7	Y8	Y0	Y7	Y8	Y0
DS0	89.4	97.8	98.3	89.2	98.2	99.0
DS1	99.3	98.4	99.3	99.8	99.3	99.8
DS2	97.2	98.9	98.0	97.4	99.4	98.2
DS3	59.3	99.0	83.8	49.5	99.7	79.8

次に，Y7，Y8，Y0 モデルを使用して，2017 年のデータ，2018 年のデータ，両方のデータを使用して評価する．また，背景の影響を見るために，ニューラルネットワークに入力したデータがオリジナルの場合 (そのまま) と，魚体部分を切り出した場合 (魚体のみ) について評価した結果を合わせて表 4 に示す．F 値は前述の 4 通りのデータ分割方法の平均値である．M/D 行は，使用したモデルと入力したデータの組み合わせを示している．入力データは，7 が 2017 年，8 が 2018 年を示す．

Y8 モデルは 2018 年データ，Y7 モデルは 2017 年データで高い F 値が得られたが，その逆の場合はほとんど判別が出来ていない．撮影方法や撮影環境が大きく変わったためと考えられる．両方のデータを使用して学習した Y0 モデルは，両方のデータに対して高い F 値を実現している．学習時に，バラエティのあるデータを与えることで多様な入力に対応できるようになる．

表 4: 各モデルとデータの組み合わせの F 値 [%]

M/D	Y8/8	Y8/7	Y7/8	Y7/7	Y0/8	Y0/7
そのまま	99.2	42.6	37.9	85.4	99.5	98.6
魚体のみ	98.8	52.2	42.0	88.1	98.9	98.8

4.3 実験 2

ニューラルネットワークによる雄雌の判別にあたり，ニューラルネットワークが人間が判別する時と同様な場所に注目しているのかを確認する．

図 4~7 は，ニューラルネットワーク VGG16 の最後の max_pooling の直前の層の出力から Gradcam[Selvaraju 16] を使用してニューラルネットワークが注目している特徴部位をヒートマップで強調表示したものである．どの図も，上側がオス下側がメスである．

図 4 は，そのままの入力画像を入力したもので，期待していた頭の形や尾びれの形に注目はしていない．特にオスでは，背景となっているマットの境界部分に着目している．データ取得時は，雄雌それぞれを連続で撮影したため，背景のマットの位置や汚れが似たようなものになるのは想像できる．ニューラルネットワークは，魚の形だけでなく背景の汚れや皺も含めて画像全体を見て判別している．

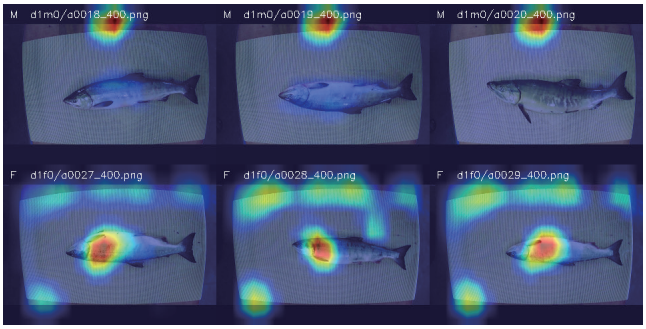

図 4: Gardcam 出力

図 5 は，入力画像から魚体部分のみを切り出した画像を入力した時のものである．図 4 と違い，背景部分への着目は見られない．メスでは図 4 で着目していた部分に同じ様に着目しているが，オスでは，ほとんど着目されなかった魚体部分に着目するようになったことがわかる．さらに，オスはエラ部分背側にメスはエラ後部腹側に着目しており，着目点が異なることがわかる．

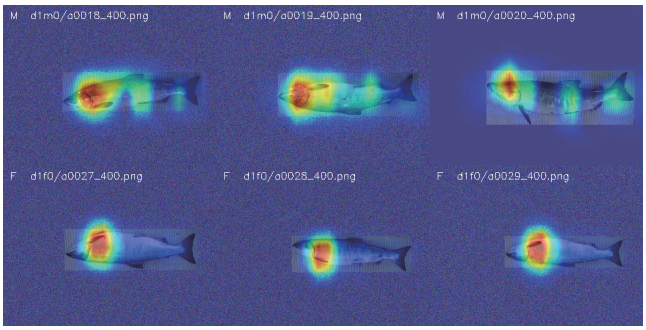

図 5: 魚体切り出し時の Gardcam 出力

人間が判断する時に着目する頭と尾の部分が，ニューラルネットワークの判断にどの程度寄与しているかを見るために，ニューラルネットワークに (1) 腹部のみを見せる，(2) 頭と尾のみを見せる，の 2 通りで評価を行った．gradcam の結果をそれぞれ図 6,7 に示す．腹部のみ見せた場合，メスは図 5 で着目していた場所と同様の場所に着目，オスは図 5 で着目していた領域は隠されたため，それ以外の領域で着目度の比較的高かった部分を着目するようになった．ニューラルネットワー

クは，部分的な情報でも，与えられた情報を最大限注視して判断しようとしていることがわかる．

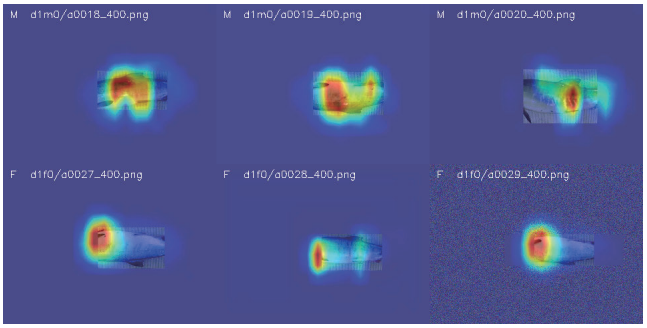

図 6: 魚体切り出し，腹部のみの場合の Gradcam 出力

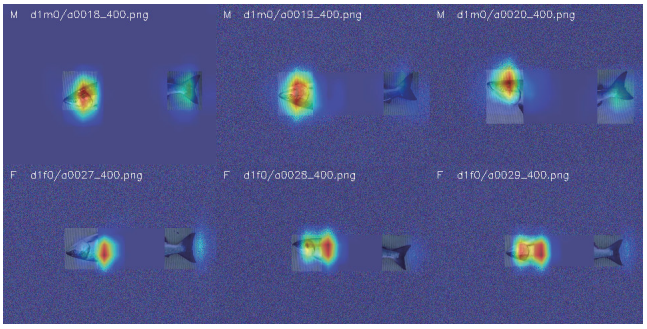

図 7: 魚体切り出し，頭と尾の場合の Gradcam の出力

表 5 に，gradcam で実験した 4 通りについての評価結果を示す．入力データをそのまま与えたものと魚体部分を切り出したものは，ほぼ同程度の精度を実現している．魚体を一部隠したもの同士を比較してみると，腹部のみを見せた方が，頭と尾を見せたものよりも精度が高い．このことからも，ニューラルネットワークは，人間が主に判断を行っている，頭と尾だけではなく，画像全体をみて判断している．画像を部分的に隠しても別の場所の情報を利用して判断することから，ロバストな判別が行えると考える．さらに，入力画像から魚体部分を全てを隠したデータを与えてみた．2018 年のデータでは，魚体を見せずに背景だけでも 80%以上の精度で判別した．学習時の DA の際に，20%の確率で背景付きの画像を与えた (表 2 の項目 1) ため，その情報も利用するようになったと考えられる．

表 5: 魚体切り出し時の各モデルでの F 値 [%]

M/D	Y8/8	Y7/7	Y0/8	Y0/7
入力そのまま	99.2	85.4	99.5	98.6
魚体のみ	98.8	88.1	98.9	98.8
魚体腹部のみ	95.5	88.7	95.5	94.3
魚体頭と尾	85.6	80.7	95.4	83.4
魚体を全て隠す	85.2	30.0	84.3	42.4

5. まとめ

ニューラルネットワークを使用したサケの雌雄判別を試行した．2017,8 年にオスメスのラベル付きサケ画像データを収集し，VGG16 のネットワークを用いて，シンプルに判別を学習させた．2017 年と 8 年で画像データの撮影方法を変えた．同一年の学習データとテストデータの組み合わせでは高い認識精度が得られた．学習データとテストデータの取得年度が異なるとほぼ認識できない結果となったが，学習に両方のデータを使用すれば，高い認識精度が得られた．高い認識精度を得るためには，魚体だけでなく背景についてもバラエティに富んだデータを得ることが重要である．

Gradcam を用いてニューラルネットワークの注視点を観察した．魚体の頭と尾のみを見せた場合，魚体の胴体部分を見せた場合，魚体を全て隠した場合について調べた．人間が判別する場合と異なり，胴体部分に着目した方が精度が高いことがわかった．魚体を隠しても 2018 年のデータでは，背景のみでも 80%以上の精度で判別できることがわかった．学習の際には，魚体部分のみを使用するなど，背景の影響を可能な限りなくす方法が望ましい．

サケ画像からの雌雄判別は微妙な差を持つ画像の 2 値分類であるが，これに対しても，ニューラルネットワークが効果的でロバストな判別を行えることを示した．人間による目視では判別できないと言われているタラやフグなどについても，ニューラルネットワークでは魚体全体の特徴を見て判別できる可能性があると考える．工業製品などの画像と違い，生物の画像では背景をコントロールすることが難しい場合が多く，データ取得の条件が悪い．このような状況で取得した画像データで学習を行うためには，背景の情報を適切にコントロールすることが必須である．今後は，サケの狩猟時期の時系列変化や漁獲地の違いによる変化などに対応できるよう，データセットを充実させていきたい．

謝辞

サケ画像の取得に協力いただいた，株式会社朝倉商店，株式会社カネコメ田中水産の皆様と研究室の学生諸君に感謝します．本研究は東京工科大学 AI 研究会の助成を受けたものです．

参考文献

[Levi 15] Levi, G. and Hassner, T.: Age and gender classification using convolutional neural networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 34–42 (2015)

[Selvaraju 16] Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Batra, D.: Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization, *arXiv preprint arXiv:1610.02391*, pp. 1–24 (2016)

[Verma 14] Verma, A. and Vig, L.: Using convolutional neural networks to discover cogntively validated features for gender classification, in *Soft Computing and Machine Intelligence (ISCMI), 2014 International Conference on*, pp. 33–37IEEE (2014)

[阿部 18] 阿部 周司, 石畑 宏明 : CNN を用いたサケの雌雄判別の試行, 信学技報, 第 117-484 巻 of *IE2017-149*, pp. 177–182, 沖縄 (2018), 2018 年 3 月 9 日 (金)　沖縄 (IE)