

研究論文

# 水産資源管理に向けた Mask R-CNNによる Few-shot 魚体長認識

長谷川 達人<sup>1,a)</sup> 田中 基貴<sup>1</sup>

受付日 2021年10月10日, 採録日 2022年2月10日

**概要：**2018年に本国において70年ぶりの漁業法の一部改正が決定された。新漁業法の主な目的は、水産資源の持続的な利用を確保し漁業生産力を発展させることとされている。この目的の達成には、操業や水揚げ等の情報を用いた資源調査を行い、科学的な知見に基づいた客観的な指標により資源評価を行い、評価結果に基づく指標に従った資源管理を行うことが重要である。一方、漁獲した尾数や、魚種、魚体長といった資源評価に必要な基礎情報は、各漁場で手動で計測されていることが多い。本研究では、Mask R-CNNを用いた画像認識により漁獲物の基礎情報を自動で収集するシステムを提案する。特に、Copy-Paste Augmentation (CP-Aug) と敵対的訓練により少量のラベル付きデータのみを用いてモデルを訓練した点と、クラス分類 Head を 1 class 分類に変更した点が特色である。評価実験の結果、CP-Augにより複数魚の検出精度が大幅に向上した。また、敵対的訓練によりアノテーション誤差に頑健な特徴表現の獲得がなされ、1 class 分類により未知魚に対する精度向上も達成した。

**キーワード：**水産資源管理, Few-shot 魚体長認識, Mask R-CNN

## Few-shot Fish Length Recognition by Mask R-CNN for Fisheries Resource Management

TATSUHITO HASEGAWA<sup>1,a)</sup> MOTOKI TANAKA<sup>1</sup>

Received: October 10, 2021, Accepted: February 10, 2022

**Abstract:** In 2018, the partial revision of the Fishery Act for the first time in 70 years was decided in Japan. The main purpose of the new Fisheries Act is to ensure sustainable use of fishery resources and to develop fishery productivity. In order to achieve this objective, it is important to conduct resource surveys using information on operations and landings, to evaluate resources using objective indicators based on scientific knowledge, and to manage resources according to the indicators based on the evaluation results. However, the basic information required for resource evaluation, such as the number of fish caught, fish species, and fish length, is often measured manually at each fishing ground. In this study, we propose a system that automatically collects basic information of fish catches by image recognition using Mask R-CNN. In particular, we trained the model using only a small number of labeled data by Copy-Paste Augmentation (CP-Aug) and adversarial training, and changed the fish species classification head to one class classifier. As a result of the evaluation experiments, the segmentation accuracy of multiple fishes was greatly improved by CP-Aug. In addition, the adversarial training resulted in the acquisition of feature representations that are robust to annotation errors, and the 1-class classification also improved the accuracy for unknown fish.

**Keywords:** fisheries resource management, few-shot fish length recognition, Mask R-CNN

### 1. はじめに

本国は世界有数の広大な漁場に囲まれており水産業の国際競争力が高く、水産業は国民に対して安定的長期的な水

<sup>1</sup> 福井大学大学院工学研究科  
Graduate School of Engineering, University of Fukui, Fukui  
910-8507, Japan  
<sup>a)</sup> t-hase@u-fukui.ac.jp

産物を供給する立場でも重要な役割を担っている。世界の漁業が安定的な生産を維持している状況に対して、国内では近年漁業・養殖業生産量が減少傾向にあり [1], 水産庁では「水産政策の改革」が進められている。限りある資源を有効かつ効率的に利用し持続可能な水産業の実現に向けた業務改革, および技術改革が求められている。

安定的長期的な水産資源の確保に向け、本国では 2018 年に 70 年ぶりの漁業法改正が決定された。漁業法<sup>\*1</sup>では、水産資源の保存や管理等に関する基本的制度を定めることで、水産資源の持続的な利用の確保や水面の総合的な利用を図り、漁業生産力を発展させることが目的であると定められている。この目的の達成に向けた改正の要点 [2] の 1 つに「新たな資源管理システムの構築」がある。

水産業における資源管理とは、最大持続生産量を実現できる資源水準を維持するために、最新の科学的知見に基づいて漁獲可能量を設定し、水産資源を管理することである。そのため、資源管理は以下の実施手順を繰り返すことにより構成される。

**資源調査** 対象の分布や回遊状況に応じて調査船や市場で漁獲情報や海洋情報、漁獲物の魚体長、体重等の基礎情報を調査する。

**資源評価** 資源調査の結果をもとに漁業や資源の現状分析を行い、資源の診断や生物学的許容漁業量 (ABC) の算定を行う。

**資源管理** ABC に基づき漁獲可能量 (TAC) を個別漁獲割当 (IQ) 方式により漁船ごとに割り当て、漁獲量の管理を行う。

**漁獲** 各漁業者は TAC に基づいて操業を行う。

上記の資源管理手順には運用上まだまだ課題が残っている。依田ら [3] や平松ら [4] の研究では、資源管理における ABC 算定の時間遅れによる影響を指摘している。たとえば 2021 年の ABC は 2020 年度の資源評価に基づいており、資源評価には 2019 年までのデータが用いられる。このようにデータの収集から指標の算出までのタイムラグによって TAC 算定に悪影響を及ぼす可能性が指摘されている。また、西田 [5] も資源評価における資源量等の推定精度向上の必要性を述べており、そのためにはより正確かつ詳細な漁獲物の基礎情報 (漁獲量や魚体長等) が不可欠となる。また、資源評価の対象魚種も 2018 年度 50 種であったところから、2023 年度までに段階的に 200 種程度にまで拡大することとなっている。すなわち、資源調査の高速化、高精度化、詳細化が求められている状況である。

資源調査は国立研究開発法人水産研究・教育機構および各都道府県の水産試験場の職員らが手作業で行うことが多



図 1 職員らによる資源調査の様子

Fig. 1 State of resource survey by staffs.

い。たとえば市場における漁獲物の体長等調査では図 1 に示すように、水揚げからせりまでの短い時間内において、職員らが手作業で魚体長の測定と記録を行っている。限られた時間内での手作業による測定では、測定できる尾数に限界があるためサンプル数が少なくなり、母集団の推定精度低下につながる。漁業従事者の人手不足も深刻であり調査のために漁業運用を妨害はできない。すなわち、資源調査の高度化を進めつつも、労力の削減を行うことも重要となる。

以上をふまえ、本研究では資源調査の一部である市場での漁獲物の基礎情報 (魚種、尾数、魚体長、魚体高) を高速かつ正確に自動収集するシステムの実現を目指す。特に本稿では、魚体長の推定に向け少量の訓練データでの魚領域認識手法を提案する。提案手法は、ベルトコンベアを用いて漁獲物を運搬、仕分けしている市場を対象として、小型計算機とカメラを搭載した機材をコンベアをまたぐように設置するだけで、コンベアを流れる漁獲物 1 尾 1 尾を画像認識により検出する。少量のラベル付きデータのみを用いて訓練できるよう工夫された Mask R-CNN モデルを用いて認識機能を実現する。最終的に、導入設置の手間やコストが少なく、かつ資源管理の課題である資源調査の高速化、高精度化、詳細化に向けた魚領域検出手法および魚体長認識手法を実現することを本研究の目的とする。

## 2. 関連研究

### 2.1 漁獲物の自動検出

Tseng らの研究 [6] では、船上を撮影した画像から釣り上げた大型魚種を検出する手法を提案している。検出には ResNet101 で入力画像を特徴量に変換し、Feature Pyramid Network (FPN) [7] を用いて Mask R-CNN [8] により実現されている。訓練データにはブリやその他を含め 3,200 枚の画像から 3,525 オブジェクトをラベル付けして用いている。Allken らの研究 [9] では、海中で撮影した魚画像の魚種を Convolutional Neural Network (CNN) により推定する手法を提案している。ImageNet で事前訓練された Inception-v3 [10] に Global Average Pooling 層と Dropout 層を追加して過学習を抑止したモデルを採用している。特

<sup>\*1</sup> 漁業法 (昭和二十四年法律第二百六十七号), 施行日: 令和二年十二月一日 (平成三十年法律第九十五号による改正), <https://elaws.e-gov.go.jp/document?lawid=324AC0000000267> (参照 2021-07-20)

に画像処理アプリケーションを用いて画像中の魚領域を手動で切り出し、背景のみを撮影した画像にランダムに貼付けを行うことでデータ拡張を行っている点が特徴である。

Álvarez-Ellacuría ら [11] は、箱に整列されたヨーロッパアンヘイクを対象とした魚体長推定手法を提案している。箱の上部から撮影された画像から Mask R-CNN により魚の頭部をインスタンスセグメンテーションし、頭部のサイズを推定する。推定した頭部のサイズから統計的手法に基づいて魚の全長を推定している。Deep Vision 社のトロール船に向けたプロダクト\*2では、トロール網に取り付けた専用機器で遊泳中の魚を撮影し、魚を引き上げなくても漁獲量を制限する方法を提供している。Garcia らの研究 [12] では Deep Vision を用いて撮影したステレオ画像から魚のインスタンスセグメンテーションとサイズ推定を行っている。前処理として照明の調整を行い、Mask R-CNN で魚領域を検出し、勾配ベースの境界推定手法を用いて重なりのある魚の検出精度向上も図っている。最終的にモルフォロジー変換とカーブフィッティングにより魚のサイズを推定している。訓練データには 1,805 枚の画像から 2,629 オブジェクトをラベル付けして用いている。

## 2.2 本研究の立ち位置

以上をふまえ、本研究の立ち位置は少量のラベル付き訓練データを用いて、ベルトコンベア上を流れる漁獲物の基礎情報を認識するシステムを実現する点である。上述の従来研究では、いずれも 3,000 オブジェクト程度のアノテーションを要することから、日本各地、世界各地に広がる漁場に展開するには大変な労力を要する。特に、水産業においては、地域ごと、季節ごとに漁獲物の種や見た目、大きさが変わることも珍しくないことから、様々な場所や条件にロバストな画像認識技術が不可欠となる。そこで本研究では、少量のラベル付き訓練データを用いるだけで実現可能な画像認識システムの実現により、様々な漁場に容易に展開可能な認識基盤の実現を目指す。少量のデータからモデルを訓練するタスクを Few-shot 学習と呼び、タスクの難易度を N-way-K-shot で表現する。ここで N は分類クラス数、K は各クラスの訓練データ数である。Few-shot 学習は一般に N-way-1-shot や 5-shot を対象とする [13]。本研究では 5 魚種を対象に各魚種 5 枚（または 1 枚）の訓練データを用いる 5-way-5-shot (1-shot) の物体検出タスクに取り組む。ベースとなるモデルは関連研究と同様に Mask R-CNN を採用する。Few-shot 学習が実現できるような工夫として、画像合成によるデータ拡張を行うことや、画像合成か否かを敵対的に弁別する Head を挿入すること、魚種分類を 1 class 分類器に変更することの 3 つの改善手法を導入し、認識モデルを実装した点が技術的な新規性である。

したがって、本研究の最終目標である漁獲物の基礎情報の認識のうち、本稿では魚体長認識に焦点を当て議論する。

## 3. 漁獲物の自動検出手法

### 3.1 システム概要

提案する漁獲物の自動検出手法を搭載した AI モニタのシステム概要を図 2 に示す。提案手法はベルトコンベア上を流れる漁獲物を対象に上部からカメラで動画を撮影し、画像中の各漁獲物の領域を検出し、魚体長を推定する。AI モニタのハードウェアは特に限定されないが、動画を撮影するカメラ、一時保存のための記憶装置、データ転送を行うための通信装置を最低要件とする。撮影環境によっては照明機器を要する。現状は撮影した動画をサーバに転送しサーバ側で画像認識を行うことを想定しているため、撮影機材はスマートフォン等でも構わない。

サーバ側に搭載される画像認識手法は図 3 に示すように Mask R-CNN [8] をベースに漁獲物のサイズ検出に特化した 3 点の工夫を行ったモデルである。1 つ目は小規模なラベル付きデータのみを用いて訓練するために Copy Paste Augmentation [14] (以降、CP-Aug と呼ぶ) を用いたデータ拡張を行った点である。2 つ目は CP-Aug を用いたことによる新しい損失関数 ( $L_{cp-adv}$ ) を提案したという点である。3 つ目は魚という抽象クラス情報を用いたセグメンテーションを行うために魚種の推定を度外視して 1 Class 検出問題とした点である。以降の章で、各手法に関して詳細に述べる。

### 3.2 Model Architecture

#### 3.2.1 Mask R-CNN

Mask R-CNN [8] は Faster R-CNN [15] を拡張することでインスタンスセグメンテーションを行うモデルであり、関連研究で示したように多くの研究で活用されている手法である。ピクセル単位でクラスを予測するセマンティックセグメンテーションタスクとは異なり、インスタンスセグメンテーションでは同一クラスの中でもインスタンスを区別する必要がある。そこで、U-Net [16] のように各ピクセルのクラスを予測するような構造ではなく、Mask R-CNN

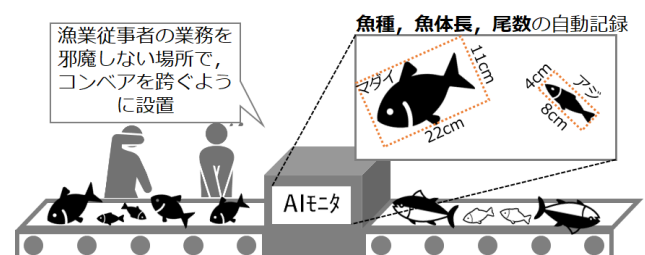


図 2 提案する漁獲物の自動認識手法

Fig. 2 Our proposed automatic fish caught recognition method.

\*2 Deep Vision, <https://deepvision.no/deep-vision/deep-vision>



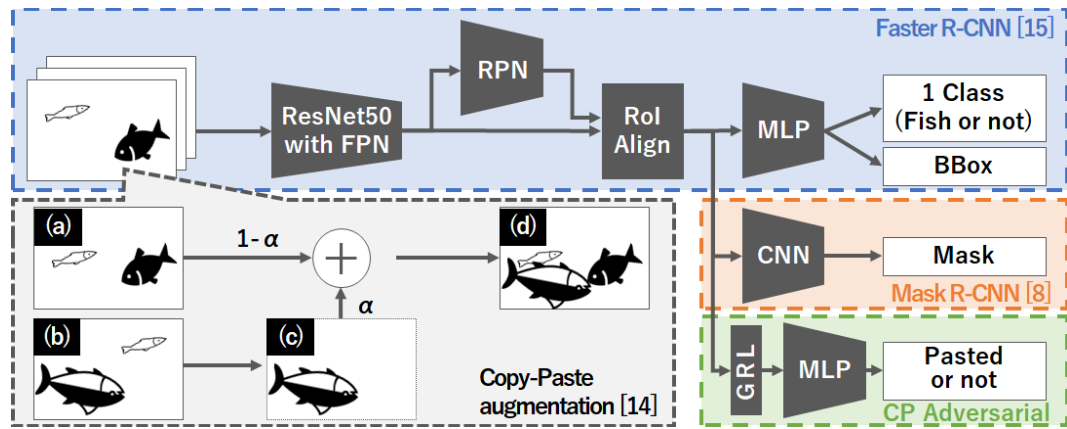


図 3 提案する魚体長自動認識手法のモデル構造

Fig. 3 Model architecture of our proposed automatic fish length recognition method.

は検出対象を囲う矩形領域である Bounding Box (bbox) を予測しつつインスタンスのマスクを予測するアーキテクチャを採用している。

図 3 の最上段が Faster R-CNN である。画像を特徴マップに変換する CNN モデルである backbone を経て、得られた特徴マップを Region Proposal Network (RPN) に入力して物体の候補領域を獲得する。今回 backbone には、関連研究同様 FPN 付きの ResNet50 を採用した。続いて、RoI Align によって、backbone で得られた特徴マップから RPN で得られた候補領域を切り出す。切り出された各候補領域の特徴マップは、続く Multi Layer Perceptron (MLP) を経てクラスと bbox の予測を行う Head に接続される。クラスは cross-entropy loss で、bbox は smooth-L1 loss で損失関数が計算されモデルを最適化する。ここで、図 3 中の FPN [7] は backbone モデルが畳み込みとプーリング処理により段階的に入力画像を縮めていく際に、各フェーズでの特徴マップをそれぞれ出力する機能である。これにより様々なスケールの対象物の検出をサポートする。

図 3 の右中段に示すように、Mask R-CNN は Faster R-CNN の RoI Align 以降に Head を追加することで、マルチタスク学習により各 bbox に対するマスクを予測する仕組みを実現している。RoI Align が出力する各候補領域の特徴マップをさらに CNN に入力することによって bbox 内のマスクの予測を実現する。マスクはクラスごとに binary cross-entropy loss によって損失関数が計算される。

### 3.3 小規模訓練データ対策

#### 3.3.1 Copy Paste Augmentation (CP-Aug)

社会実装を行う際に重要な課題となるのが訓練データのアノテーションである。本研究では Mask R-CNN を扱うため、画像内の各魚に対して魚種ラベルを付与し、魚領域のセグメンテーションマスクを作成する必要がある。魚 1 尾に対して 2 分程度アノテーションに時間を要すると考えると、関連研究のように 3,000 オブジェクトのアノテ

ションに 100 時間を要することとなる。また、漁場が変わるにともない訓練データを作り直す手間を考えると、アノテーションはなるべく手間を少なく抑えられることが望ましい。

前述の Allken らの研究 [9] では、手動で魚領域を切り出し背景のみの画像に貼り付けることでデータ拡張を行っていた。同様にセグメンテーションタスクに対して、Ghiasi らがコピー&ペーストによるデータ拡張手法 CP-Aug を提案している [14]。CP-Aug では図 3 左下に示すように、訓練時のミニバッチ内からランダムに Target 画像 (a) と Source 画像 (b) を選択し、Target 画像に対して Source 画像中のランダムなオブジェクト (c) をコピー&ペーストすることでデータ拡張を行う (d)。このとき、該当の魚種ラベルとマスクラベルもコピー&ペーストする点が特徴である。合成画像は  $I_1 \times \alpha + I_2 \times (1 - \alpha)$  で算出される。ここで  $I_1$  は Source 画像、 $I_2$  は Target 画像、 $\alpha$  は Source 画像のバイナリマスクである。なお、 $\alpha$  に対しガウシアンフィルタを用いることで、2 画像間の境界をぼかした自然な合成を実現できる。本稿ではデフォルト設定を踏襲し  $\sigma = 1.0$  のガウシアンフィルタを  $\alpha$  に適用した。

#### 3.3.2 CP-Adversarial

本研究では CP-Aug を採用したこととともなって、Mask R-CNN に挿入する新たな Head として CP-Adversarial Head を提案する。CP-Aug を用いる際、合成画像ならではの周囲のピクセルとの不調和が発生しうる。これによってマスク検出 Head が不調和を判断根拠としてマスクを検出するように、Backbone モデルが特徴表現の獲得を目指す可能性がある。そこで図 3 右下に示すように、Gradient Reversal Layer (GRL) [17] と MLP を経て CP-Aug によってペーストされたオブジェクトなのか否かを推定する CP-Adversarial Head を新たに追加する。GRL は順伝播時には恒等関数として働くが、逆伝播時には勾配の正負を反転する Layer である。GRL により敵対的訓練モデルを単一の最小化問題として表現することが可能となる。こ

のとき、CP-Adversarial Head の損失関数  $L_{cp-adv}$  は式 (1) で定義される。

$$L_{cp-adv} = -\frac{1}{M} \sum_{i=1}^M \{y_i \cdot \log z_i + (1-y_i) \cdot \log(1-z_i)\} \quad (1)$$

ここで、 $M$  はミニバッチのサイズ、 $y_i$  は CP-Aug によってペーストされたオブジェクトか否かを示す正解ラベル、 $z_i$  は CP-Adversarial Head の MLP の出力に対し Sigmoid 関数を適用した値である。すなわち、式 (1) は CP-Aug によってかさ増しされたオブジェクトか否かを識別する binary cross-entropy loss である。

CP-Adversarial Head は GRL により逆伝播時に勾配の正負が反転する。そのため、MLP 部の訓練は画像合成を見破るように訓練され、backbone は画像合成を見破れないように訓練される。これによって、CP-Aug による周囲ピクセルとの不調和を判断材料としないような特徴表現の獲得を促す効果が期待できる。

### 3.3.3 1 Class Detection

Mask R-CNN は各クラスが独立であることを仮定して設計されたモデルである。たとえば、犬や机のようにそれぞれ独立したカテゴリに属する物体を検出するため、図 3 右上の分類器では、本来検出対象の物体数を認識するための多クラス交差エントロピー誤差を用いて最適化がなされる。たとえば、ブリ、サバ、アジといった多クラス分類を行うモデルとして設計する場合、それぞれが独立して学習されるため魚という共通する特徴表現の獲得を困難にする可能性がある。

関連研究として、Li ら [18] は物体検出タスクにおいて抽象クラスの情報を併用する手法を提案している。クラス分類器を抽象クラスの分類器と元のクラス分類器で分岐させマルチタスク学習を行う手法となっている。

上記をふまえ本研究では、図 3 右上に示すクラス分類器を「魚」か「否」かのみを分類する 1 Class 分類器として扱う。これによって、将来的に様々な漁場をまたいだ運用を行う際に、新規魚種であっても追加訓練なしで魚として検出することが可能となる。また、マスクラベルをすべてのデータにアノテーションすることは非常に膨大な手間がかかるが、魚種のラベルを付与することは比較的容易である。したがって魚種の検出は別モデルとして分離することで、魚種の検出と魚体長の検出を実現する。

## 4. 評価実験

### 4.1 実験設定

本研究で提案する漁獲物の魚体長認識手法に関して、実験により認識精度評価を行う。本稿では魚種の判別は度外視し、インスタンスセグメンテーションとしてのマスク検出精度と、魚体長の推定精度に関して検証を行う。

#### 4.1.1 データ収集

2020 年 6 月に水産研究・教育機構と佐賀県唐津魚市場の協力のもと、照明環境の整った屋内で魚を撮影することで本研究で使用するデータセットを収集した。三脚で高さ約 1m に固定したデジタルカメラ (Nikon 製 COOLPIX W300) を用いて、カゴ内に魚を手動でランダムな位置に配置し、直上からフラッシュを焚きつつ撮影を行った。ベルトコンベア環境を擬似的に再現したものであるため画角が変わらぬよう配慮しているが、三脚のずれや日をまたぐことによるずれが多少ある環境である。同一の 1 尾につき座標と角度を変え 5 枚～20 枚程度撮影している。撮影対象はサバ、カマス、アジ、アイゴ、ムツの 5 種 (図 4) である。実験で用いた詳細な画像数は次項で説明する。撮影は 4K 画質で行ったが本実験では縦 864, 横 1,152 pixel に縮小して分析を行った。

#### 4.1.2 データセットの作成

撮影した画像データに対して手動で魚種ラベルとマスクラベルを付与しデータセットを作成した。本実験で使用するデータセットの詳細を表 1 に示す。まず、1 枚の画像中に複数の魚が写った全 81 枚の画像データをテスト用データ (複数) とする (以下、 $D_{test}^{multi}$  と呼ぶ)。 $D_{test}^{multi}$  は表 1 の魚種内訳で、のべ 340 尾の魚が写っているが、撮影日の都合上アイゴの画像のみ含まれていない。次に、1 枚の画像中に 1 尾の魚が写った画像から各魚種 50 尾ずつをランダムに選定しテスト用データ (単体) とした (以下、 $D_{test}^{single}$  と呼ぶ)。同様に、1 枚の画像中に 1 尾の魚が写った画像のうち  $D_{test}^{single}$  を除いた画像から各魚種 K 尾ずつ、計 5K 枚を

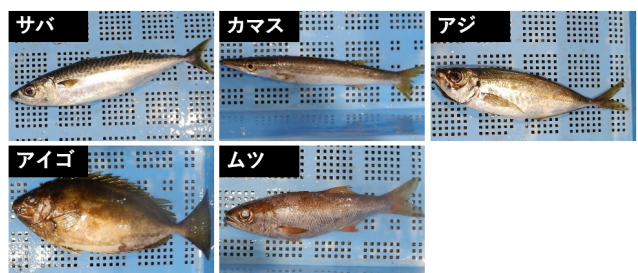


図 4 撮影対象とした 5 魚種 (サバ、カマス、アジ、アイゴ、ムツ)  
Fig. 4 5 fish species of our target photograph (mackerel, sphyrænaidae, horse mackerel, siganus fuscus, gnomefish).

表 1 データセットの内訳 (K は shot 数であり、5-shot の場合は 25 枚の画像を  $D_{train}$  とする)

Table 1 Details of the dataset (K is a number of K-shot. 5-shot means  $D_{train}$  includes 25 pictures).

	画像枚数	撮影尾数				
		サバ	カマス	アジ	アイゴ	ムツ
$D_{train}$	5K	K	K	K	K	K
$D_{test}^{single}$	250	50	50	50	50	50
$D_{test}^{multi}$	81	80	81	116	0	63

表 2 魚の領域検出に関する Ablation study の結果 (1 class 分類の mAP)

Table 2 Results of ablation study about fish segmentation (mAP for 1 class detection).

Method				$D_{test}^{single}$		$D_{test}^{multi}$	
Name	Copy-paste	Ext-head	Adversarial	1 class	5 class	1 class	5 class
None	-	-	-	0.838 ± 0.002	0.759 ± 0.005	0.380 ± 0.028	0.262 ± 0.014
CP-Aug	o	-	-	<b>0.841</b> ± 0.001	0.815 ± 0.002	<b>0.782</b> ± 0.004	0.699 ± 0.019
CP-Adv	o	o	o	0.839 ± 0.001	0.820 ± 0.003	0.777 ± 0.004	0.700 ± 0.017
CP-Cls	o	o	-	0.822 ± 0.004	0.761 ± 0.004	0.755 ± 0.005	0.575 ± 0.006

ランダムに選定し訓練用データとした (以下,  $D_{train}$  と呼ぶ). 以降, 特に明記されていない場合は  $K=5$  で  $D_{train}$  が 25 枚の画像であるものとする.

アノテーションは手作業で実施され, すべての画像には魚種とマスクラベルが付与されている. なお,  $K=5$  の  $D_{train}$  のアノテーションにおおよそ  $2 \text{分} \times 25 \text{尾} = 50 \text{分}$  の時間を要している.  $D_{test}$  のアノテーションにおおよそ  $2 \text{分} \times 590 \text{尾} = 1,180 \text{分}$  として, 計 19 時間半程度の時間を要している. なお, 実運用時には本来テスト用データ  $D_{test}$  は不要なため, 各漁場に展開する際には K-shot に応じて 10K 分程度のアノテーションを要することを想定している.

#### 4.1.3 モデルの実装と訓練設定

実験で使用するモデルは Pytorch の torchvision を用いて実装した. torchvision では Mask R-CNN が提供されているため, これに独自の CP-Adversarial Head を追加する実装を行った. backbone は ResNet50 に FPN を併用したモデルとし, ImageNet の事前訓練による重みを利用している. CP-Aug は github の実装例<sup>\*3</sup>をもとに CP-Adversarial Head 用の拡張を行った.

モデルの訓練は 400 エポック行い, Stochastic Gradient Descent (SGD) を学習率 = 0.005, momentum = 0.9, weight decay = 0.0005, learning scheduler で訓練の 1/3 ごとに学習率を 0.1 倍する設定で最適化した. 最適化に用いた各パラメータは探索的に決定した. データ拡張は画像認識のデータ拡張ライブラリ Alubmentations [19] より, 上下左右反転, 上下左右シフト, スケール変換, 回転をランダムに適用している. なお, 25 枚で構成される  $D_{train}$  で提案手法を 400 エポックの訓練するには, Nvidia RTX 3090 を搭載した計算機を用いて 1 時間程度の訓練時間を要する.

#### 4.1.4 評価指標

評価指標は COCO データセットの物体検知用 Metrics<sup>\*4</sup>から mAP を採用した. AP は Average Precision である. Intersection over Union (IoU) が一定以上を検出成功とみなして, 予測確信度を降順に並べた際に Recall が 1.0 となるまでの Precision を積分した値である.

特に mAP は 0.5 から 0.95 まで 0.05 刻みで IoU 基準値を変化させた際の平均 AP である. その他の指標として AR (Average Recall) があるが, 今回 AR は AP と同様の傾向を示したこともあり付録で議論する.

## 4.2 セグメンテーション精度の評価

### 4.2.1 Ablation study

提案手法の有効性を検証するための Ablation study を行った結果を表 2 に示す. 表中の結果は 5 試行の mAP の平均 ± 標準偏差を示している. 縦方向はモデルの種類を示しており, それぞれ Copy-paste の有無, 追加 Head (Ext-Head) の有無, 敵対的訓練 (Adversarial) の有無を示している. 最下段の Ext-Head あり Adversarial なしは, Ablation のために実装したモデル (CP-Cls) であり, 図 3 の GRL を取り除くことで画像合成の有無をマルチタスク学習するモデルである. 横方向にはテストデータの種類 ( $D_{test}^{single}$ ,  $D_{test}^{multi}$ ) と, クラス分類器を 1 class として訓練したか, 5 class として訓練したかを示している. なお 5 class で訓練した場合も評価時には 1 class 問題として mAP を算出している.

モデルの種類に着目すると,  $D_{test}^{single}$ ,  $D_{test}^{multi}$  いずれも CP-Aug が最良, 次いで CP-Adv と, 画像合成による mAP の向上が確認された. 特に, 単一魚種の検出問題  $D_{test}^{single}$  に対しては, None と比べ mAP が 0.003 程度の差であったが, 複数魚の検出問題  $D_{test}^{multi}$  では mAP が 0.402 と大幅に向上する結果となった. 1 枚の画像に 1 尾が写った画像のみの訓練データセットからは, 魚の重なり等の影響を学習することができなかったことが要因であると考えられる.

CP-Aug による精度向上が確認された一方で, CP-Adv を導入したことで精度向上には至らぬ結果となり, mAP をわずかに低下させる結果となった. しかしながら, CP-Cls に着目すると, CP-Aug に対して 2~3%程度の mAP 低下が確認された. すなわち, 画像合成を弁別する損失関数を付加することで CP-Cls は画像合成による不調和を過度に特徴として抽出させた可能性がある. これに対し, CP-Adv は mAP 向上には至らなかったものの, 不調和によらない特徴表現の獲得が実現できたと考えられる.

1 class 分類による影響を考察すると, いずれの条件においても 1 class 分類として訓練したモデルのほうが mAP

<sup>\*3</sup> conradry/copy-paste-aug, <https://github.com/conradry/copy-paste-aug>

<sup>\*4</sup> COCO Detection Evaluation, <https://cocodataset.org/#detection-eval>



表 3 マスクラベルを収縮 (−10 px, −5 px), 膨張 (+10 px, +20 px) させた際のセグメンテーション精度評価 (1 class 分類の mAP; 5 試行)

Table 3 Segmentation accuracy evaluation (mAP for 1 class detection; 5 trials) when the mask label is shrunk (−10 px, −5 px) or is expanded (+10 px, +20 px).

		Mask transform pixels of training data				
		−10	−5	0	+10	+20
$D_{test}^{single}$	None	0.440 ± 0.003	0.738 ± 0.001	0.838 ± 0.002	<b>0.569</b> ± 0.006	<b>0.307</b> ± 0.007
	CP-Aug	<b>0.639</b> ± 0.006	<b>0.817</b> ± 0.001	<b>0.841</b> ± 0.001	0.545 ± 0.001	0.237 ± 0.003
	CP-Adv	0.633 ± 0.004	0.815 ± 0.001	0.839 ± 0.001	0.541 ± 0.002	0.236 ± 0.003
	CP-Cls	0.519 ± 0.010	0.780 ± 0.002	0.822 ± 0.004	0.552 ± 0.006	0.227 ± 0.004
$D_{test}^{multi}$	None	0.229 ± 0.004	0.366 ± 0.019	0.380 ± 0.028	0.223 ± 0.008	0.096 ± 0.003
	CP-Aug	0.589 ± 0.003	<b>0.764</b> ± 0.002	<b>0.782</b> ± 0.004	<b>0.504</b> ± 0.003	0.208 ± 0.001
	CP-Adv	<b>0.596</b> ± 0.003	0.760 ± 0.004	0.777 ± 0.004	0.500 ± 0.003	<b>0.211</b> ± 0.008
	CP-Cls	0.454 ± 0.007	0.703 ± 0.006	0.755 ± 0.005	0.468 ± 0.009	0.178 ± 0.007

が高い結果となった。したがって、魚か否かを推定するモデルとしたことで、魚に共通する特徴表現の獲得につながり、セグメンテーション精度が向上したと考えられる。

#### 4.2.2 訓練データのマスクラベルが不正確な際の影響調査

CP-Aug はマスクに基づいてコピー&ペーストを行う手法であるため、マスクが正確にアノテーションされていることが前提である。一方、アノテーションは人間が手動で行うため、マスクには多少のノイズが付与される可能性がある。そこで、マスクにノイズが付与された環境を擬似的に再現して影響の調査を行った。正確にアノテーションされたマスクに対して、モルフォロジー変換により領域を収縮 or 膨張させることで、擬似的に不正確なマスクを再現した。収縮は行いすぎるとマスクがなくなってしまうため、収縮の幅は −10 px と −5 px を、膨張の幅は +10 px と +20 px を採用した。

表 3 に各モデル 5 試行した際の結果を示す。表中の横軸はマスクを収縮 or 膨張させた度合いを、縦軸はモデルの種類を示している。当然ながらマスクの収縮 or 膨張が大きくなるほど mAP が低下する。 $D_{test}^{single}$  に対しては、マスク収縮時は CP-Aug と CP-Adv が有効に働く一方で、マスク膨張時は None が最良となった。マスクを膨張させて合成することで、Source 画像中に写る魚を隠してしまうことが原因であると考えられる。 $D_{test}^{multi}$  に対しては、全体的に CP-Aug と CP-Adv が同程度に良い mAP を示しており、特に収縮や膨張が顕著なとき (−10 や +20) には CP-Adv の方がわずかに高い mAP となった。このことから、敵対的訓練を導入することにより画像合成による不調和の影響をやや緩和できていることを確認した。

#### 4.2.3 未知の魚に対する影響調査

1 class 分類モデルによる、訓練時に未知の魚に対する領域検出精度を評価するため、我々の先行研究 [20] で提案した Blender を用いたデータ拡張手法により合成画像データセット  $D'_{train}$  を作成した。各魚種 1 枚の画像からランダムに配置された複数魚の合成画像データ 3,000 枚を自動

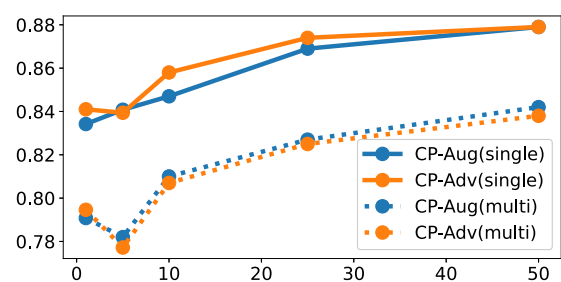


図 5 訓練データ数を変化させた際のセグメンテーション精度変化 (1 class 分類の mAP)

Fig. 5 Change of segmentation accuracy by the number of training samples (mAP of 1 class detection).

生成している。ただし、本データにはアイゴが含まれていない画像のみを採用している。これに対して、単一のアイゴのみを含むテスト用データセット  $D'_{test}$  を作成し、訓練データに含まれない未知の魚に対するセグメンテーション精度評価実験を行った。図 4 に示したとおり、アイゴは今回対象 5 魚種の中でも、他と形状や色相が大きく異なる魚種であったため、未知魚として採用した。

検証の結果、5 class 分類とした場合には、mAP が 0.486 であったが、1 class 分類とした場合には mAP が 0.854 と、0.368 の大幅な mAP 向上が見られた。このことから、1 class 分類としてモデルを訓練することで魚全体に共通する特徴表現の獲得を促し、未知の魚に対しても正確にセグメンテーションが行えることを確認した。

#### 4.2.4 訓練データ数の変化の考察

ここまでの考察は訓練データ数が各魚種 5 尾ずつの画像で与えられる前提で実施してきた。ここで、訓練データ数を増減させた際の mAP の変化を図 5 に示す。横軸は訓練データに含まれる各魚種の画像枚数 (K-shot における K) であり、縦軸は mAP である。なお、K の最小は 1 としたが、K = 5 に対してデータサイズが 1/5 となったことから、K = 1 の場合のみ訓練エポック数を 400 から 2,000 に変更した。

図 5 より、訓練データが増加するにともなって mAP も

表 5 魚体長の推定誤差 (平均  $\pm$  S.D.) [mm]Table 5 Estimation error of fish length (Avg.  $\pm$  S.D.) [mm].

		bbox		MaskA		MaskB	
		CP-Aug	CP-Adv	CP-Aug	CP-Adv	CP-Aug	CP-Adv
補正なし	全長	12.17 $\pm$ 7.72	12.12 $\pm$ 7.91	5.66 $\pm$ 4.72	6.16 $\pm$ 5.04	<b>5.37</b> $\pm$ 4.22	5.50 $\pm$ 4.45
	尾叉長	29.24 $\pm$ 10.09	29.04 $\pm$ 10.35	17.03 $\pm$ 8.36	<b>16.70</b> $\pm$ 8.76	21.06 $\pm$ 7.43	20.91 $\pm$ 7.65
補正あり	全長	6.95 $\pm$ 5.60	7.13 $\pm$ 5.92	5.60 $\pm$ 4.65	6.10 $\pm$ 4.89	<b>4.65</b> $\pm$ 3.90	4.93 $\pm$ 4.12
	尾叉長	8.05 $\pm$ 6.09	8.25 $\pm$ 6.25	7.17 $\pm$ 5.22	7.57 $\pm$ 5.39	<b>5.98</b> $\pm$ 4.46	6.28 $\pm$ 4.48

増加するが、増加幅を見るといずれのケースも  $K = 1$  の mAP に対して、 $K = 50$  で 0.030 程度の増加となっていることが確認できる。mAP 差 0.030 の良し悪しに関しては、次の節の魚体長の推定精度で議論する。

### 4.3 魚体長推定の精度検証

#### 4.3.1 実験概要

本節では魚体長の検出精度に関する議論を行う。前節の  $K = 1$  の場合における 2 モデル (CP-Aug, CP-Adv) を用いて、 $D_{test}^{single}$  に対する魚体長推定を行う。

ここで、魚体長には複数の計測方法がある。上顎の先端から尾鰭の付け根までの長さを「標準体長」、そこからさらに尾鰭が二又する中央部の凹みまでの長さを「尾叉長」、さらに尾鰭の先端までの長さを「全長」と呼ぶ。尾鰭は回遊中に欠損することが少なくないため、資源調査では尾叉長を計測することが一般的である。本稿では、データセットの特性上「全長」と「尾叉長」をベースに議論する。今回準備したマスクラベルはすべて尾の先までを含んでいるため、提案手法は「全長」が推定できる。なお、マスクラベルを尾鰭が二又する中央部の凹みまでに意図的に制限することで、提案手法は「尾叉長」を推定することもできる。

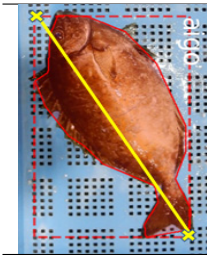
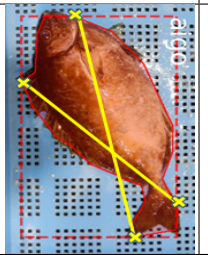
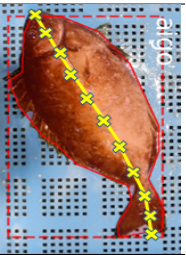
精度評価にあたって、正確な魚体長ラベルが必要となる。本稿の撮影実験では漁港の業務を妨害しないため、実際の漁獲物の長さを計測することができていない。そこで、今回はすべてのテスト画像上で手動で長さを計測し、pixel 値を mm 単位に変換することで正解データとした。正解ラベルおよび、推定結果の各算出値 (pixel) はハイパーパラメータとして与えられたカゴの底辺サイズ (mm) を用いた比率で mm 単位に変換される。

なお、今回の撮影画像には撮影機材の都合上、歪曲収差が起り、かご上の漁獲物の配置によっては端に寄った漁獲物で最大 5 mm 程度の誤差が発生する可能性がある。したがって、実際の運用時には歪曲収差の補正が可能な機材を用いるか、撮影後から提案手法利用前までの前処理として、歪曲収差の補正を行うことを前提とする。

提案手法はバウンディングボックス (bbox) とセグメンテーションマスクを推定可能なため、これを魚体長に変換する必要がある。本稿では簡易的に表 4 に示す 3 手法で各ラベルから全長 (pixel) に変換した。各手法の詳細は付

表 4 各ラベルからの魚体長算出手法

Table 4 Method to calculate fish length from each label.

bbox	MaskA	MaskB
		
bbox の対角線を採用する方式	マスクから上下左右の頂点を交差する線分の長い方を採用する方式	マスクから物体の中心をトレースする方式 (詳細は付録参照)

録で説明する。

#### 4.3.2 実験結果

提案手法により推定を行った平均絶対誤差 (MAE) を表 5 に示す。MAE は  $MAE = \frac{1}{N} \sum_{i=1}^N |e(i)|$  で算出される。ここで、 $N$  はデータ数、 $i$  番目のデータに対する誤差は  $e(i) = d_t(i) - d_p(i)$  であり、正解ラベル  $d_t(i)$  と推定値  $d_p(i)$  から算出される。これに対し、尾叉長や MaskB による推定値は正解値よりも長めになる傾向があることから、補正パラメータ  $c$  を減じて  $MAE_c = \frac{1}{N} \sum_{i=1}^N |e(i) - c|$  で補正ありの推定誤差を算出した。今回補正パラメータ  $c$  は既知であると仮定し  $c = \frac{1}{N} \sum_{i=1}^N |e(i)|$  とした。

表 5 より、補正パラメータが未知の場合、CP-Aug と MaskB による手法で全長推定の MAE が 5.37 mm となった。bbox や MaskA による手法は MaskB にやや劣る精度となった。今回のデータセットでは尾の先までを検出するため、尾叉長推定の MAE は CP-Adv の MaskA で 16.70 mm となった。これに対し補正パラメータが既知の場合、全長で  $MAE = 4.65$  mm、尾叉長で 5.98 mm となった。

表 6 に CP-Aug ( $K = 1$ ) による魚種ごとの補正ありの全長推定誤差を示す。すべての魚種で MaskB による手法が最良精度となり、推定精度の高い魚種としてムツが  $MAE = 3.43$  mm、低い魚種としてアイゴが  $MAE = 6.05$  mm となった。表中の平均全長より、カマスやアイゴのような比較的全長の長い魚種では MAE が増加する傾向が見られた。

なお、実際の資源調査現場においては 5 mm 刻みで魚体



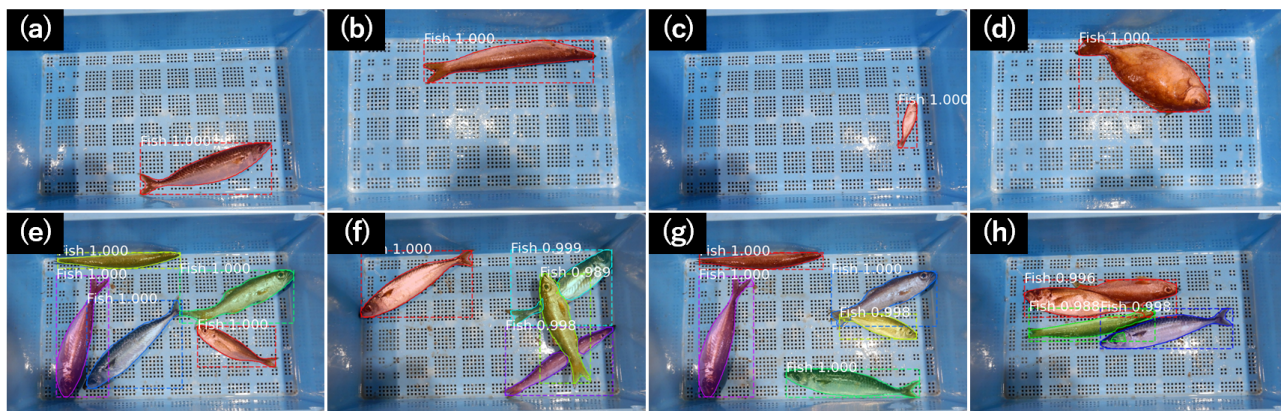


図 6 インスタンスセグメンテーション実施事例 (K = 50 の CP-Aug)

Fig. 6 Instance segmentation sample results (CP-Aug with K = 50).

表 6 CP-Aug (K = 1) による魚種ごとの補正ありの全長推定誤差 (平均 ± S.D.) [mm]

Table 6 Corrected estimation error of total length by CP-Aug (K = 1) for each fish species (Avg. ± S.D.) [mm].

名称	全長	MaskA	MaskB
サバ	272.00 ± 31.07	4.99 ± 3.77	<b>4.48</b> ± 3.56
カマス	304.69 ± 35.50	6.74 ± 5.32	<b>5.78</b> ± 4.05
アジ	188.34 ± 58.74	3.97 ± 2.85	<b>3.49</b> ± 2.64
アイゴ	270.17 ± 53.06	7.50 ± 5.95	<b>6.05</b> ± 5.20
ムツ	244.33 ± 10.96	4.77 ± 3.68	<b>3.43</b> ± 2.65

長の計測を行っていることから、平均絶対誤差 4.65 mm は実運用に耐えうる誤差範囲であると考えている。

#### 4.4 セグメンテーションの定性評価

最後に、図 6 でインスタンスセグメンテーションの成功事例と失敗事例を定性的に評価する。各マスクは本実験で最良精度を示した K = 50 時の CP-Aug を用いて推定した。図上部には  $D_{test}^{single}$  に対するセグメンテーション結果である。特に 1 枚の画像に 1 尾の魚が写っているケースでは、ほぼ誤差なく魚領域の検出ができていていることが分かる。特に (c) は小アジであり、他の魚と大きくサイズが異なるが正確に検出できている。

図下部は  $D_{test}^{multi}$  に対するセグメンテーション結果である。1 枚の画像に複数魚が写るケースでは、(e) のように大部分は正確にマスクが検出できる一方で、(h) のように魚が密集している際に複数魚をまとめて 1 尾と検出してしまうケースがいくつか見られた。しかし、(f) のように魚が完全に上下に重なってしまっているケースを検出できている場合もあり、提案手法の検出力の高さがうかがえる。

以上をふまえ、定性的にマスクを観察すると、全体的に高精度にマスクが検出できていることが確認できた。そのうえで、補正ありでも 4.65 mm の全長推定誤差が発生しているため、マスク検出後の後処理によるマスクから魚体長への変換手法は改善の余地がある。

## 5. おわりに

本研究では、水産業における資源管理の改善に向けて、資源調査における漁獲物の正確な魚体長認識手法の開発を行った。提案手法はベルトコンベア上を流れる漁獲物を上部からカメラで撮影し、Mask R-CNN ベースの認識手法により漁獲物のサイズを検出する。少量のラベル付きデータを用いた Few-shot な訓練手法を提案し精度評価を行った。実験の結果、1 class 分類とした Mask R-CNN を CP-Aug で訓練することで、5-way-1-shot の訓練データでも、4.65 mm 程度の誤差で魚体長の検出が行えることを確認した。また、1 class 分類として魚種分類を別問題とすることで mAP が向上することや、新たに提案した CP-Adversarial Head の導入により、画像合成に依存しない特徴表現の獲得が行えることを確認した。一方で、定性評価では密集した魚の切り分けが難しいという課題が明らかとなったことから、今後は密集した魚の検出精度を向上させるベースモデルの検討や、マスクから魚体長に変換する後処理手法の改善を行っていく。

謝辞 本研究は、JST ACT-X (JPMJAX20AJ) の支援を受けたものであり、ここに感謝の意を表する。なお、本研究は水産研究・教育機構開発調査センターの行う海洋水産資源開発事業（定置網：佐賀県玄海地区）から得られたデータの一部を含み、ご協力をいただいた同センターの皆様、佐賀玄海漁協の皆様感謝の意を表する。また、資源調査の現場視察や、現場運用の議論にご協力いただいた福岡県水産試験場の皆様に感謝の意を表する。

#### 参考文献

- [1] 水産庁：令和 2 年度 水産白書，第一部 水産の動向，特集 マーケットインの発想で水産業の成長産業化を目指す，第 1 節 国内外の水産物マーケットをめぐる状況，(2) 国内の水産物需給をめぐる状況 (2018)，入手先 (<https://www.jfa.maff.go.jp/j/kikaku/wpaper/r02.h/trend/1/t1.f1.2.html>) (参照 2021-12-03)。

- [2] 水産庁：平成 30 年度 水産白書，水産の動向，第二部 水産施策，1 水産政策の改革，2 「漁業法等の一部を改正する等の法律」について (2018)，入手先 ([https://www.jfa.maff.go.jp/j/kikaku/wpaper/h30\\_h/trend/2/t2\\_01\\_2.html](https://www.jfa.maff.go.jp/j/kikaku/wpaper/h30_h/trend/2/t2_01_2.html)) (参照 2021-12-03)。
- [3] 依田真里，渡邊千夏子，由上龍嗣，福若雅章：資源管理スケジュールが管理効果に及ぼす影響，日本水産学会誌，Vol.82，No.5，pp.676–685 (オンライン)，DOI: 10.2331/suisan.15-00075 (2016)。
- [4] 平松一彦，土田優斗：生物学的許容漁獲量 (ABC) 算定における時間遅れの資源管理への影響，日本水産学会誌，Vol.advpub (オンライン)，DOI: 10.2331/suisan.20-00031 (2020)。
- [5] 西田 宏：新たな資源管理システムの構築 (資源評価・資源管理) について，日本水産学会誌，Vol.86，No.5，pp.434–434 (オンライン)，DOI: 10.2331/suisan.WA2674-2 (2020)。
- [6] Tseng, C.-H. and Kuo, Y.-F.: Detecting and counting harvested fish and identifying fish types in electronic monitoring system videos using deep convolutional neural networks, *ICES Journal of Marine Science*, Vol.77, No.4, pp.1367–1378 (online), DOI: 10.1093/icesjms/fsaa076 (2020)。
- [7] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S.: Feature Pyramid Networks for Object Detection, *Proc. CVPR 2017*, pp.936–944 (online), DOI: 10.1109/CVPR.2017.106 (2017)。
- [8] He, K., Gkioxari, G., Dollár, P. and Girshick, R.: Mask R-CNN, *2017 IEEE International Conference on Computer Vision (ICCV)*, pp.2980–2988 (online), DOI: 10.1109/ICCV.2017.322 (2017)。
- [9] Allken, V., Handegard, N.O., Rosen, S., Schreyeck, T., Mahiout, T. and Malde, K.: Fish species identification using a convolutional neural network trained on synthetic data, *ICES Journal of Marine Science*, Vol.76, No.1, pp.342–349 (online), DOI: 10.1093/icesjms/fsy147 (2018)。
- [10] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z.: Rethinking the Inception Architecture for Computer Vision, *Proc. CVPR 2016*, pp.2818–2826 (online), DOI: 10.1109/CVPR.2016.308 (2016)。
- [11] Álvarez-Ellacuría, A., Palmer, M., Catalán, I.A. and Lisani, J.-L.: Image-based, unsupervised estimation of fish size from commercial landings using deep learning, *ICES Journal of Marine Science*, Vol.77, No.4, pp.1330–1339 (online), DOI: 10.1093/icesjms/fsz216 (2019)。
- [12] Garcia, R., Prados, R., Quintana, J., Tempelaar, A., Gracias, N., Rosen, S., Vågstøl, H. and Løvall, K.: Automatic segmentation of fish using deep learning with application to fish size measurement, *ICES Journal of Marine Science*, Vol.77, No.4, pp.1354–1366 (online), DOI: 10.1093/icesjms/fsz186 (2019)。
- [13] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H. and Hospedales, T.M.: Learning to Compare: Relation Network for Few-Shot Learning, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018)。
- [14] Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E.D., Le, Q.V. and Zoph, B.: Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation, *Proc. CVPR 2021*, pp.2918–2928 (2021)。
- [15] Ren, S., He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R. (Eds.), *Proc. NIPS 2015*, Vol.28 (2015)。

- [16] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, *Proc. MICCAI 2015*, pp.234–241 (2015)。
- [17] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M. and Lempitsky, V.: Domain-Adversarial Training of Neural Networks, *Journal of Machine Learning Research*, Vol.17, No.59, pp.1–35 (2016) (online), available from (<http://jmlr.org/papers/v17/15-239.html>)。
- [18] Li, K., Wang, N.Y., Yang, Y. and Wang, G.: SGNet: A Super-class Guided Network for Image Classification and Object Detection, *Proc. CRV 2021*, pp.127–134 (online), DOI: 10.1109/CRV52889.2021.00025 (2021)。
- [19] Buslaev, A., Parinov, A., Khvedchenya, E., Iglovikov, V.I. and Kalinin, A.A.: Albumentations: Fast and flexible image augmentations, *ArXiv e-prints* (2018)。
- [20] 田中基貴，近藤和真，増田花乃，長谷川達人：オブジェクト自動生成による魚用インスタンスセグメンテーションのデータ拡張手法，人工知能学会全国大会論文集，Vol.JSAI2021，p.3G4GS2i02 (オンライン)，DOI: 10.11517/pjsai.JSAI2021.0.3G4GS2i02 (2021)。

## 付 録

### A.1 Recall に関する評価

将来的な尾数カウントを考慮したとき，評価指標として Recall を用いた議論が必要となる．領域検出の Recall 指標として，Average Recall (AR) という指標がある．AR は予測確信度を降順に並べ上位  $n$  件を採用した際の Recall である．今回  $n = 10$  とした  $AR^{max=10}$  を評価指標とし，Ablation Study を行った結果を表 A.1 に示す．なお，全体の傾向は mAP と同じである．

4.4 節の定性評価で，複数尾の魚が 1 カ所に密集している際に，複数尾の魚をまとめて 1 尾と検出されるケースがあった． $K = 1$  の CP-Aug の  $D_{test}^{multi}$  に対する予測結果を目視で確認した結果，340 尾のうち 2 尾ずつ 3 組が 1 尾として認識されていることを確認した．多くの画像は図 6 の (e) や (g) のように，個体間に間が空いていたためうまく検出できていたが，(h) のように間なく密集していた場合において，比較的多く誤認識が確認された．

### A.2 各ラベルから全長への変換

表 4 に示した 3 手法による各ラベルから全長 (pixel) への変換手法について詳細に説明する．

表 A.1 魚の領域検出に関する Ablation study の結果 (1 class 分類の  $AR^{max=10}$ )

Table A.1 Results of ablation study about fish segmentation ( $AR^{max=10}$  for 1 class detection).

Method	$D_{test}^{single}$	$D_{test}^{multi}$
None	$0.857 \pm 0.001$	$0.528 \pm 0.014$
CP-Aug	<b><math>0.863 \pm 0.001</math></b>	<b><math>0.812 \pm 0.004</math></b>
CP-Adv	<u><math>0.860 \pm 0.001</math></u>	<u><math>0.809 \pm 0.002</math></u>
CP-Cls	$0.846 \pm 0.003$	$0.789 \pm 0.004$

**A.2.1 bbox**

検出されたインスタンスを囲む bbox の対角線の長さを全長として扱う方式である．表 4 のように漁獲物が斜めに配置される場合，正確な予測が可能となるが，水平，垂直に配置される場合，誤差が大きくなる．

**A.2.2 MaskA**

4 方の bbox に隣接する接点を 4 点算出し，上下左右に交差するように線分を結んだ際に最長となる線分を採用する方式である．表 4 の例では上下を結ぶ線分のほうがやや長いのでこれを採用している．

**A.2.3 MaskB**

マスクから物体の中心を検出し，トレースする方式である．まず，マスクと接する画素が多い対角線を選択する．次に，対角線の各画素における垂線とマスクが重なる画素の始点と終点を検出する．始点と終点を結ぶ線分の中点を記録する．対角線のすべての画素から抽出した中点系列を，対角線を水平にする回転角  $\theta$  で回転し，最小二乗法により 3 次関数にフィッティングする．フィッティング結果を  $-\theta$  で回転した結果から，各中点系列間のユークリッド距離の合計値を全長とする．



長谷川 達人 (正会員)

2009 年石川工業高等専門学校電子情報工学科卒業．2011 年金沢大学工学部情報システム工学科卒業．同年株式会社富士通北陸システムズ入社．2014 年東京医療保健大学助手．2015 年金沢大学大学院自然科学研究科修了．博士（工学）．2017 年福井大学大学院工学研究科講師を経て 2020 年より同大学准教授．専門はコンテキストウェアネスと教育工学．IEEE 会員．



田中 基貴

2021 年福井大学工学部電気電子情報工学科卒業．同年より同大学大学院工学研究科知識社会基礎工学専攻博士前期課程在学中．