

水産資源管理に向けた Mask Keypoint R-CNNによる魚体長認識

長谷川 達人^{1,a)} 田中 基貴¹

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 Keypoint 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、漁獲物認識のためのKeypointを新たに定義した点、Copy-Paste Augmentationにより少数のラベル付きデータのみを用いてモデルを訓練した点が特色である。本稿では、資源管理のための基礎情報として利用できる精度で自動収集が実現可能かという観点で提案手法を評価した結果を報告する。

キーワード：水産資源管理, 漁獲物認識, Keypoint 検出

Fish Length Recognition by Mask Keypoint R-CNN for Fisheries Resource Management

TATSUHITO HASEGAWA^{1,a)} MOTOKI TANAKA¹

Abstract: In 2018, the partial revision of the Fisheries Law 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, evaluate resources using objective indicators based on scientific knowledge, and manage resources according to the indicators based on the evaluation results. However, the basic information required for resource assessment, such as the number of fish caught, fish species, and fish length, is often measured manually at each fishing port. In this study, we propose a system that automatically collects basic information of catches by image recognition using Mask Keypoint R-CNN. In particular, we define a new keypoint for catch recognition, and train the model using only a small number of labeled data by Copy-Paste Augmentation. In this paper, we report the results of the evaluation of the proposed method from the viewpoint of whether it is feasible to collect the data automatically with accuracy that can be used as basic information for resource management.

Keywords: Fisheries Resource Management, Fish Catch Recognition, Keypoint Detection

1. はじめに

2018年の水産庁水産白書によると水産業は世界的にも成

長産業であり、全世界の漁業生産量は年々増加傾向にある。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産物を供給する立場でも重要な役割を担っている。一方、国内では近年漁獲量が減少傾向にあり、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

¹ 福井大学大学院工学研究科
Graduate School of Engineering, University of Fukui
^{a)} t-hase@u-fukui.ac.jp

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、及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^{*1}では、水産資源の保存や管理など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 [1] の一つに「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 [2] や平松ら [3] 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。例えば 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 [4] 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・教育機構及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い。例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



図 1 職員らによる資源調査の様子

Fig. 1 State of resource survey by staffs

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。

以上を踏まえ、本研究では資源調査の一部である市場での漁獲物の詳細情報 (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を開発する。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 Keypoint 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 [5] 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 を backbone とし、Feature Pyramid Network (FPN)[6] を用いて Mask R-CNN[7] により実現されている。訓練データにはブイやその他を含め 3200 枚の画像から 3525 オブジェクトをラベル付けして用いている。Allken らの研究 [8] 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[9] に Global Average Pooling 層と Dropout 層をを追加して過学習を抑止したモデルを採用している。特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 [10] 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

^{*1} 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=324AC00000000267> (2021.7.20 Accessed)

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、魚を引き上げずとも漁獲量を制限する方法を提供している。Garcia らの研究 [11] 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 1805 枚の画像から 2629 オブジェクトをラベル付けて用いている。

2.2 本研究の立ち位置

以上を踏まえ、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、ベルトコンベア状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 3000 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、ごく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。画像認識のベースは関連研究と同様に、ResNet50 と FPN を backbone とした Mask R-CNN を採用する。これに正確な魚体長検出のための Keypoint 検出を組み込むこと、少量のラベル付き訓練データで訓練が実現できるような工夫としてコピペによるデータ拡張や、新しい損失関数を組み込む手法を開発する点が技術的な新規性である。

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画像を撮影し、画像中の各漁獲物の尾数、種類、魚体長、魚体高を推定する。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[7] をベースに漁獲物のサイズ検出に特化した 4 点の工夫を行ったモデルである。1 つ目は魚の検出に特化した Keypoint を定義し Keypoint 検出 Head を併用した点である。2 つ目は魚という Super Class 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。3 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[12] (以降、CP-Aug と呼ぶ) を用いたデータ拡張を行った点である。4 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。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 Keypoint R-CNN

Mask R-CNN[7] は Faster R-CNN[13] をシンプルに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[14] 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 は Bounding Box (bbox) を予測しつつインスタンスの Mask 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。画像を入力として backbone モデルに与え、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。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全体を最適化する。ここで、図 3 中の FPN[6] 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

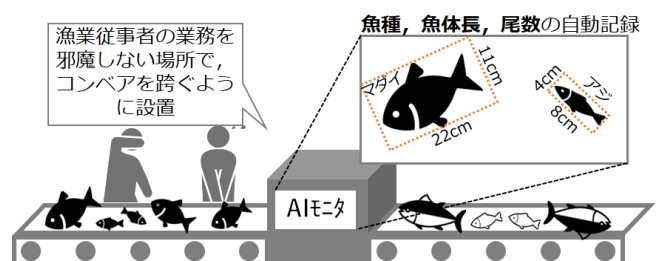


図 2 提案する漁獲物の自動認識手法

Fig. 2 Our proposed automatic fish caught recognition method.

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

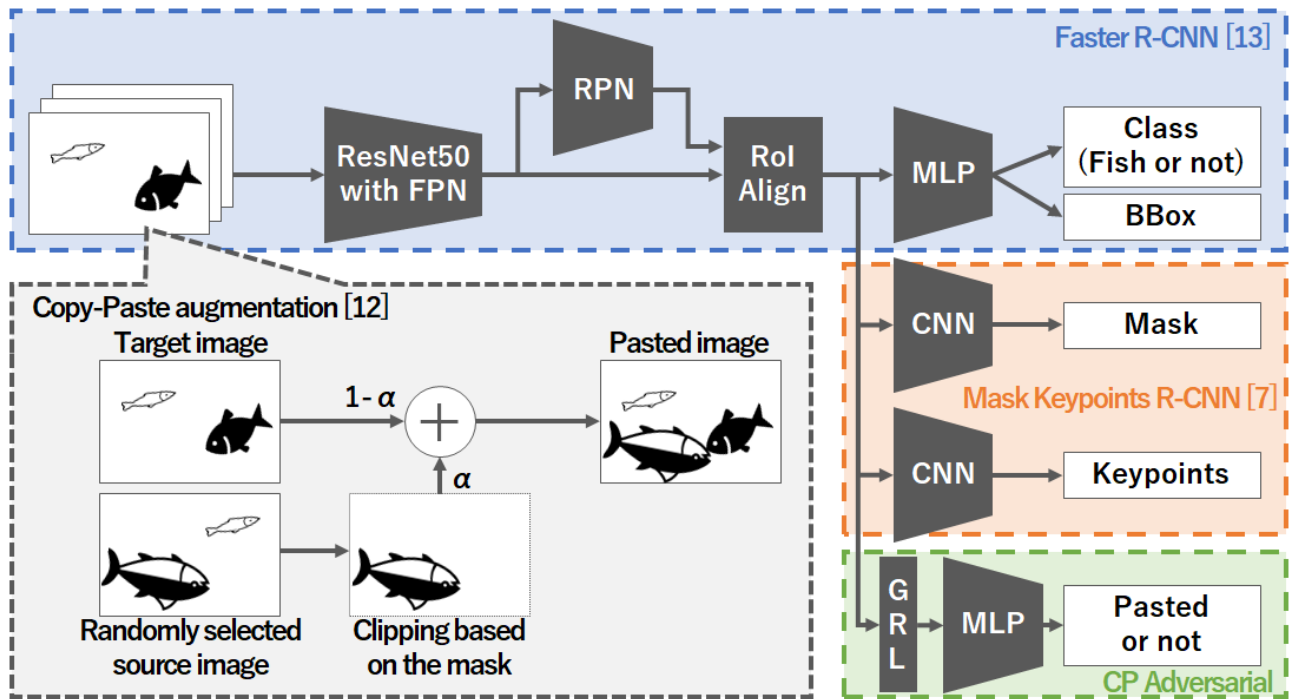


図 3 提案する自動認識手法のモデル構造

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

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 Mask や Keypoints 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更に CNN に入力することによって bbox 内の Mask や Keypoints の予測を実現する。Mask はクラスごとに各マスク値に関する binary cross-entropy loss によって損失関数が計算される。Keypoints は K 種の Keypoint に対して One-hot ベクトル形式の Mask を割り当て、Keypoint に該当するセルが 1 となるように cross-entropy loss で損失関数が計算される。

3.2.2 漁獲物のサイズ認識のための Keypoints 定義

これまで深層学習を用いた Keypoints 検出は様々な分野で応用されているが、魚の Keypoints 検出を行った事例は筆者の調べた限りでは見つからない。そこで本研究では、福井県水産試験場職員の方々との議論の上で、漁獲物のサイズ認識に向けた Keypoints を図 4 のように 9 点定義した。それぞれ、口 (Mouth)、目 (Eye)、背びれ (Dorsal fin)、胸びれ (Pectoral fin)、腹びれ (Pelvic fin)、尾 (Tail root, Tail top, Tail center, Tail bottom) と一般的な魚が備えている各部位とした。特に資源調査における魚体長は図中の Mouth から Tail root にかけての標準体長 (Standard length) として定義されることから Keypoints の検出はサイズ推定に役立つと考えられる。

3.3 小規模訓練データ対策

3.3.1 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。例えば、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、Class 認識 Head では cross-entropy loss を用いて最適化がなされる。したがって、本研究における Class 認識 Head を、例えばブリ、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共通する魚という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 [15] は物体検出タスクにおいて Super class の情報を併用する手法を提案している。Class 認識 Head を Super class の分類器と実際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。

上記を踏まえ本研究では、図 3 に示す Class 認識 Head を魚か否かのみを分類する 1 Class Detection モデルとして扱う。これによって、将来的に様々な漁場をまたいだ運用

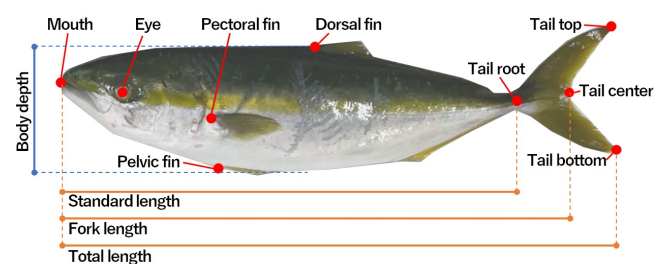


図 4 漁獲物のサイズ認識のための Keypoints 定義

Fig. 4 Definition of keypoints for fish length recognition.

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、Mask や Keypoints の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。なお、本稿では魚体長の検出に焦点を当てて議論を行う。

3.3.2 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 Keypoint 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 Mask を作成し、9 点の Keypoints を設定する必要がある。魚 1 尾に対して 3 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 3000 オブジェクトのアノテーションに 150 時間を要することとなる。また、漁場が変わるに伴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たいという背景がある。

前述の Allken らの研究 [8] 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、Ghiassi らがコピー＆ペーストによるシンプルなデータ拡張手法 CP-Aug を提案している [12]。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 と Source 画像を選択し、Target 画像に対して Source 画像中のランダムなオブジェクトをシンプルにコピー＆ペーストすることでデータ拡張を行う。なお、ハイパーパラメータとして重み α を設定することで、合成時に Source 画像を透過することが可能だが、本稿では $\alpha = 1$ とした。

3.3.3 CP-Adversarial

本研究では CP-Aug を採用したことに伴い新たな Head として、CP-Adversarial Head を提案する。CP-Aug を用いる際、周囲のピクセルとの不調和が発生しうる。これによって Mask 検出 Head が不調和を判断根拠として Mask 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) [16] 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シンプルな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。

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

ここで、 N 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 によってかき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 CP-Aug を見破るように訓練され、backbone は CP-Aug 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せると考えられる。

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 Mask 検出精度と、魚体長の推定精度に関して検証を行う。

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。撮影対象はアジ、カマス、ムツ、アイゴ、サバの 5 種であり、計 1200 枚程度の撮影を行った。基本的には 1 枚に 1 尾が写るようにしているが、複数尾が同時に配置されている画像も一部含まれている。図 5 は撮影した画像例である。撮影は 4K 画質で行ったが本実験では 1152:864 に縮小して分析を行った。

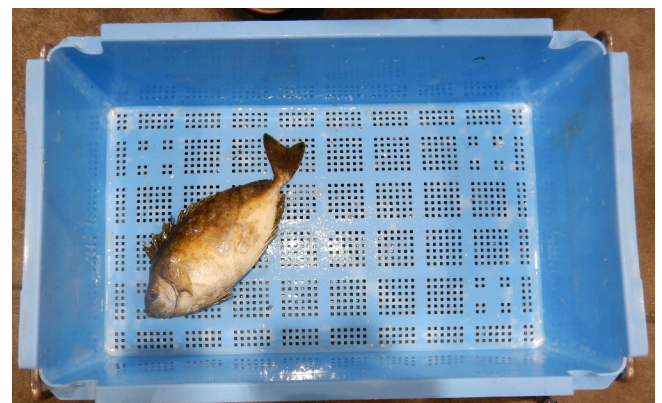


図 5 唐津魚市場における撮影画像例

Fig. 5 Sample image taken at the Karatsu fish market.

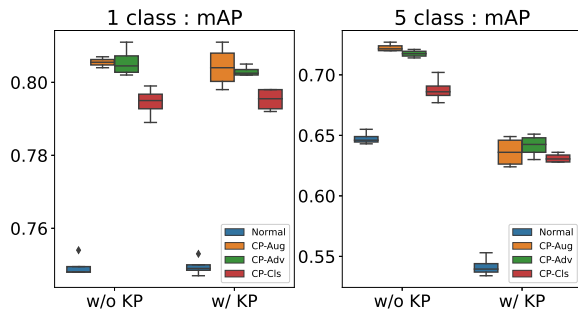


図 6 mAP の Ablation study の結果
Fig. 6 Results of ablation study (mAP).

収集したデータセットからランダムサンプリングにより訓練データ 25 枚 (以降 D_{train}), テストデータ 130 枚 (D_{test}) を抽出しアノテーションを行った。ここで, D_{train} は全て 1 枚に 1 尾のみが含まれている画像であり, 5 魚種 5 枚ずつで構成される。 D_{test} は全データセットから D_{train} を除いた画像よりランダムに 130 枚が選択され, 複数尾が同時に含まれている画像もある。アノテーションは手作業で実施され, D_{train} には魚種, Mask, Keypoints が, D_{test} には魚種, Mask のみがラベルとして付与されている。なお, D_{train} のアノテーションにおおよそ 3 分 $\times$ 25 尾 = 75 分, D_{test} のアノテーションにおおよそ 2 分 $\times$ 130 尾 = 260 分として, 計 5 時間半程度の時間を要している (なお, 実運用時には本来 D_{test} は不要なため, 各漁場に展開する際には 75 分程度のアノテーションを要することになる)。

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 と Keypoint R-CNN が独立して提供されているため, これをマージして Mask Keypoint R-CNN を独自に実装した。その上で CP-Adversarial Head も追加実装している。 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^{*3}をもとに CP-Adversarial Head 用の拡張を行った。モデルの訓練は 500 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。 CP-Aug 以外のデータ拡張は Alubmentations パッケージより Flip のみを採用した。

評価指標は COCO データセットの物体検知用 Metrics^{*4}から, mAP と $AR^{max=10}$ 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

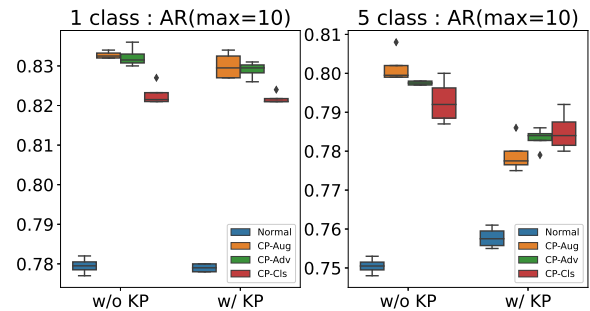


図 7 AR(max=10) の Ablation study の結果
Fig. 7 Results of ablation study of AR(max=10).

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。 AR は Average Recall であり, 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 $n=10$ の指標を採用している。

4.2 セグメンテーション精度の評価

提案手法によるセグメンテーション精度の評価実験を行った結果を図 6, 図 7 に示す。各図左は分類クラスを魚のみ (1 class) とした結果, 右は 5 魚種を推定した結果である。また, 各図は内部で Keypoint 検出なし (w/o KP) とあり (w/ KP) を表示している。各箱ひげは 5 試行の結果の集約であり, 凡例は検出手法を意味する。 Normal は通常の Mask R-CNN を, その他はデータ拡張に CP-Aug を使用した上で, CP-Adv は CP-Adversarial Head を導入した場合, CP-Cls は CP-Adversarial Head から GRL を除いた場合を意味する。すなわち, CP に対して敵対的に訓練するか協調的に訓練するかの差を見ている。

4.2.1 Keypoint 検出による精度変化

Keypoints による影響を見ると, Class によらず Keypoints 検出 Head を追加することで領域検出のスコアが僅かに低下することがわかる。今回の実験では各 Head の Loss の比率は全て均一としたことが原因で, Keypoint 検出の精度を維持するために Mask 検出精度を低下させたものと考えている。したがって, Mask 検出と Keypoints 検出は同一の backbone でマルチタスク学習を行うのではなく各々別のモデルとすることが望ましいと考えられる。

4.2.2 1 Class 分類による精度変化

分類 Class 数による影響を見ると, 当然ながら 1 Class 分類を行った方がスコアが高くなる傾向がある。各指標の計算はクラス単位で行われた上で最終的にクラス間で平均を算出するため, 領域は検出できているがクラス予測を誤った際にスコアが低下することに起因すると考えられる。

そこで, 分類 Class 数による影響を正確に比較するために, 訓練時は 5 class 分類として訓練し, 精度検証時には検

^{*3} conradry/copy-paste-aug <https://github.com/conradry/copy-paste-aug>

^{*4} COCO Detection Evaluation, <https://cocodataset.org/#detection-eval>

表 1 検出クラス数による検出精度の比較 (Mask R-CNN with CP-Aug)

Table 1 Detection accuracy comparison on the number of detection classes (Mask R-CNN with CP-Aug).

訓練時	5 class	5 class	1 class
検証時	5 class	1 class	1 class
mAP	0.7143	0.795	<u>0.8055</u>
AR(max=10)	0.8095	0.8385	0.8327

出したクラスを全て魚 (1) として検出するモデルを実装し、検出精度評価を 4 試行した結果を表 1 に示す。表より、5 class 分類としてモデルを訓練する手法よりも、最初から 1 class 分類としてモデルを訓練する手法のほうが、mAP が 1% 程度高いことが確認できる。一方、AR(max=10) は 0.6% 程度低下している。したがって、魚という Super class を有するインスタンスセグメンテーションタスクにおいて、1 class 分類でモデルを訓練し、魚種判別は別モデルで実現することで魚体長の検出で Precision 優位な推定が実現できる可能性が示唆された。

4.2.3 Copy-Paste による精度変化

Copy-Paste による影響を見ると、図の結果のいずれにおいても Normal が最も低い値を示していることから CP-Aug はマスク検出精度に大いに貢献することがわかる。次に CP-Aug と CP-Adv を比較すると、ややばらつきはあるものの概ね同程度のスコアになっていることがわかる。一方で、CP-Cls を見ると大部分で CP-Aug や CP-Adv よりもスコアが低下している。すなわち、コピペされたマスクか否かを弁別する Head を導入し協調的に訓練することで、コピペを検出できるような特徴表現を獲得するようになり、結果としてマスク検出の汎化性能を下げたと考えられる。反対に、CP-Adv によってスコアは向上しなかったものの CP-Aug と同程度となったということは、すなわち、スコア低下を抑制しつつコピペに依存しない特徴表現の獲得につながった可能性が示唆されたことになる。

なお、CP-Aug はセグメンテーション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、セグメンテーション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、領域ラベルには多少のノイズが付与される可能性がある。そこで、領域ラベル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。今回ノイズの付与は正確にアノテーションされた領域ラベルに対してモルフォロジー変換により領域を膨張させることで擬似的に再現する。すなわち、魚領域のアノテーションがずれて大きめに領域を選択してしまった場合を再現している。

図 8 に 1 class 検出を 5 試行した際の結果を示す。図中の x 軸はマスクを膨張させた度合い (0 px, 5 px, 10 px) を, y 軸はスコアである。図よりマスクの膨張に伴い CP-Adv

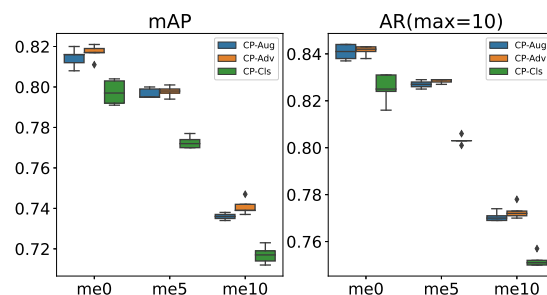


図 8 マスク拡張を行った際の各手法の検出精度 (1 class 検出, 5 試行平均)

Fig. 8 Detection accuracy for each method when the mask is expanded (average of 5 trials on 1 class detection).

の効果が顕著になる様子が確認できる。アノテーションが不正確な場合 CP-Aug によって背景の合成の違和感が増大するため、これを緩和するような特徴表現の抽出に貢献できた可能性がある。なお、mAP のマスク膨張 0 px に関して、図 6 では精度差は確認できなかったが、今回は CP-Adv の方が高いスコアとなった。これについては、実験の都合上バッチサイズを 10 から 5 に変えたことによる影響の可能性があるが、原因の追求は今後の課題としたい。

4.3 セグメンテーションの定性評価

検出結果を可視化することでインスタンスセグメンテーションの定性評価を行う．1 class 分類の CP-Aug を用いた Mask Keypoint R-CNN モデルの出力事例を用いて議論するが，予測結果をそのまま利用すると一つのインスタンスに対して複数の候補領域が出力されてしまうため，後処理が必要となる．今回，候補として得られた bbox を以下のルールにより後処理して可視化を行った．

- (1) bbox の検出スコアが 0.4 以下のものは除外する.
- (2) Non-Maximum Suppression により, Mask IoU が 0.7 以上の場合は確信度が低い方を除外する.
- (3) とあるマスクから, それ以外の残った全マスクの和集合を減じたマスクが, 元のマスクに対して 0.1 未満の比率ならば除外する (複数の魚をまたぐケースの除外).

図 9 に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示す。図上部の成功事例のように、魚と魚の間にある程度の間隔が空いている場合は、良好に領域検出ができています。また、Keypoints をみても、それぞれ検出できていない Keypoint もあるものの、概ね良好に検出ができています。一方、図下部の失敗事例のように、魚が密集している場合は複数魚をまたいで 1 尾と検出がされてしまうことで、検出に失敗している様子が確認できる。このような事例については、本研究のようにシンプルな Mask R-CNN を用いた手法では正確な検出は難しい。したがって、今後の課題として後処理の強化ないしは、重なりに強いベースモデルへの改良が必要となる。

検出成功事例



検出失敗事例



図 9 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

Fig. 9 Instance Segmentation Successes and Failures.

表 2 各ラベルからの魚体長算出手法

Table 2 Method to calculate fish length from each label.

Mask	bbox	KP	KP(Std.)
Mask から上下左右の頂点を抽出し交差するように線分を作成し、長い方を採用する方式	bbox の対角線を採用する方式	Keypoints の口、胸びれ、尾 (root)、尾 (center) をそれぞれ線分で結び、合計長を Total length として採用する方式	Keypoints の口、胸びれ、尾 (root) をそれぞれ線分で結び、合計長を Standard length として採用する方式

4.4 魚体長認識の精度評価

ここまではセグメンテーションマスクの検出精度に関して議論を行ってきたが、本節では魚体長の検出精度に関する議論を行う。前節の結果に引き続き、1 class 分類の CP-Aug を用いた Mask Keypoint R-CNN モデルの出力を用いる。なお、全 130 枚の D_{test} に写る全 147 オブジェクトのうち、適切に bbox が検出された 143 オブジェクトで議論する。除外した 4 オブジェクトは前述の通り魚の重なりにより 1 尾を適切に切り出せなかった事例である。各オブジェクトには Total length と Standard length のラベルを画像から手動で付与した。

提案手法により推定可能なラベルは Mask, bbox, Keypoints であり、これを用いて魚体長に変換処理を行う必要がある。本稿では簡易的に表 2 に示す 4 手法で各ラベルから魚体長 (px) に変換した。Mask では、4 方の bbox に隣接する接点を 4 点算出し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。図の例では左右を結ぶ線分のほうがやや長い為此を採用している。bbox では、bbox の対角線をそのまま採用する。ここで、Mask と bbox は原理的に Total length しか推定ができないことに注意されたい。KP では、検出した Keypoints のうち、口、胸びれ、尾 (root)、尾 (center) をそれぞれ線分で結び、合計長を Total

表 3 魚体長の検出誤差 (cm)

Table 3 Detection error of the fish length [cm].

	Length		Error			
	Total	Std.	Mask Total	bbox Total	KP Total	KP Std.
Mean	26.73	22.67	1.07	1.35	3.35	2.50
S.D.	5.84	5.32	1.13	0.92	4.92	4.89
Min	9.35	7.80	0.01	0.07	0.00	0.00
25%	23.48	19.57	0.33	0.65	0.62	0.24
50%	27.36	23.35	0.68	1.20	1.19	0.76
75%	29.83	25.08	1.42	1.87	4.29	1.28
Max	38.45	34.10	7.17	5.14	24.72	23.80

length として採用する。これを尾 (center) までとしたものを Standard length として採用する。各算出値 (px) 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 (cm) を用いた比率で cm 単位に変換される。

本手法により推定を行った推定誤差を表 3 に示す。左中 2 列は D_{test} 中のオブジェクトの体長の基本統計量である。右 4 列は推定値とラベルとの絶対誤差である。平均絶対誤差 (MAE; Mean 行) を見ると、Mask を用いた全長検出が最も誤差が小さく、MAE では 1cm 程度誤差で、中央値では 0.68cm の誤差で魚体長が検出できていることがわかる。KP は MAE は悪いが、Min や 25% を見ると比較的良好な検出ができていることがわかる。これは Keypoint 検出自体の精度が甘いことに起因しており、Keypoint 検出精度を向上させることで改善が可能である。特に Keypoint は予測ができなかった際に思いもよらぬ方向に出現することで、Max の誤差が非常に大きいことがわかる。Mask や bbox を併用したルールベースの手法でもこれは改善可能であると考えている。

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 Keypoint 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。25 枚という少数のラベル付きデータを用いた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 Keypoint R-CNN を CP-Aug で訓練することで、25 枚という少数の訓練データでも、概ね良好に魚の領域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 がやや向上することや、新たに提案した CP-Adversarial Head の導入により、コピー&ペースト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や後処理手法の検討を行っていく。

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、ご協力を頂いた同センターの皆様、佐賀玄海漁協の皆様に感謝の意を表する。また、魚の Keypoints 定義などにおいて議論にご協力いただいた福井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：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，https://www.jfa.maff.go.jp/j/kikaku/wpaper/h30_h/trend/2/t2_01_2.html (2021/7/20 Accessed) (2018).
- [2] 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).
- [3] 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).
- [4] 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).
- [5] 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).
- [6] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S.: Feature Pyramid Networks for Object Detection, *Proc. of the CVPR 2017*, pp. 936–944 (online), DOI: 10.1109/CVPR.2017.106 (2017).
- [7] 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).
- [8] 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).
- [9] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z.: Rethinking the Inception Architecture for Computer Vision, *Proc. of the CVPR 2016*, pp. 2818–2826 (online), DOI: 10.1109/CVPR.2016.308 (2016).
- [10] 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).
- [11] 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).
- [12] 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. of the CVPR 2021*, pp. 2918–2928 (2021).
 - [13] Ren, S., He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *Proc. of the NIPS 2015* (Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R., eds.), Vol. 28 (2015).
 - [14] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, *Proc of the MICCAI 2015*, pp. 234–241 (2015).
 - [15] Li, K., Wang, N. Y., Yang, Y. and Wang, G.: SGNet: A Super-class Guided Network for Image Classification and Object Detection, *Proc. of the CRV 2021*, pp. 127–134 (online), DOI: 10.1109/CRV52889.2021.00025 (2021).
 - [16] 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 (online), available from <http://jmlr.org/papers/v17/15-239.html> (2016).