

コンピュータ科学科
知能システム系
卒業論文

物体検出とキーポイント検出による
ヨットにおける
メインセールの形状の可視化

2024 年 2 月

102030372 宮原 光彥

要旨

ヨット競技において、ヨットはセールの内外を通過する風の気圧差による揚力を利用して推進するため、セールの形状は非常に重要な要素の一つである。そのため、選手たちは実際の海上練習を動画で撮影し、それを陸に帰ったのちに確認を行っている。しかし、動画からセールの形状を正確に把握するのは、ヨットを始めたばかりの初心者は動画を見ててもセールの形状を把握することは難しい。

本研究ではヨットの練習中に撮影された動画の1フレームから抽出した画像から、ヨットのセールの輪郭、および風を受けてはらんでいるメインセールの内側面の可視化を実現することを目的とした。

提案する手法として、物体検出手法にはYOLOv5、キーポイント検出手法にはMask R-CNNを使用した。ヨットが映る画像に対しYOLOv5にてメインセールの外接矩形を検出し、その領域を抽出した。その後、その矩形領域に対して、メインセールの輪郭上に存在する特徴点をMask R-CNNによりキーポイントとして検出した。検出したキーポイントを元に、メインセールの輪郭を可視化しメインセールの内面の着色を行った。

実際のヨットの練習動画を用いた評価実験を行い、YOLOv5による物体検出精度とMask R-CNNによるキーポイント検出精度を評価した。またMask R-CNNに関しては、複数の学習条件にて学習を行い、そのモデルのキーポイント検出精度を比較評価することで、複数のモデルの優劣を求めた。比較には有意水準5%でのF検定・t検定やモデルの物体検出数、メインセールの内面とその真値とのIoUを用いた。

YOLOv5による物体検出の精度は、Recall値が0.91と高い精度を得た。Mask R-CNNによるキーポイント検出については、最も優れているとされた学習条件は、メインセールの外接矩形とインスタンスセグメンテーション、キーポイントの検出を学習し、その際に学習用データセットのデータ拡張を行うものであった。そのモデルにて、検出されたキーポイント座標の真値との差の平均値は実世界においてx軸方向(右を正)には-7.66cmから0.89cm、y軸方向(下を正)には-27.57cmから48.90cmの差を持った。また、検出されたキーポイントを用いた内面の着色領域と真値とのIoUは平均で0.78であった。

今後の課題としては、キーポイント検出精度や輪郭抽出精度の向上が挙げられる。また、提案手法に基づいてメインセールの3D深さ推定に発展することが考えられる。

目次

第 1 章	序論	2
1.1	研究の背景と目的	2
1.2	関連研究	3
1.3	本研究の特徴と本論文の概要	3
第 2 章	本研究に用いる基本的手法と知識	5
2.1	ヨットの知識とセールの役割	5
2.2	メインセールの幾何形状	7
2.3	物体検出手法 (YOLOv5)	8
2.4	キーポイント検出手法 (Mask R-CNN[8])	12
2.5	検定方法	15
第 3 章	メインセールの形状の可視化システム	17
3.1	提案システムの概要	17
3.2	YOLOv5 による物体検出	18
3.3	Mask R-CNN によるキーポイント検出	18
3.4	メインセールの形状可視化システム	20
3.4.1	輪郭線の推定アルゴリズム	20
3.4.2	内側面の推定アルゴリズム	20
第 4 章	実験	22
4.1	YOLOv5 によるメインセール領域の検出	22
4.1.1	方法	22
4.1.2	結果	24
4.1.3	考察	25
4.2	Mask R-CNN によるキーポイントの検出	26

4.2.1	実験方法	26
	データセット	26
	ネットワーク構成	29
	評価方法	31
4.2.2	実験 1：インスタンスセグメンテーション学習の有りと無しの比較	32
	比較するモデル	32
	結果	33
	考察	35
4.2.3	実験 2：マスト，ブーム，リーチを 1 ネットワークで学習するか 3 ネットワークで学習するかの比較	36
	比較するモデル	36
	結果	37
	考察	42
4.2.4	実験 3：物体領域についてメインセールとマスト，ブーム，リーチとの比較	43
	比較するモデル	43
	結果	43
	考察	47
4.2.5	データ拡張学習に対する検定	47
	比較するモデル	47
	結果	47
	考察	49
4.3	メインセールの形状可視化の IoU 比較実験	50
4.3.1	実験方法	50
4.3.2	結果	51
	考察	52
4.4	実験結果のまとめ	53
第 5 章	結論	56
5.1	まとめ	56
5.2	今後の課題	57
参考文献		59

第1章

序論

1.1 研究の背景と目的

ヨット競技において、セールの形状は非常に重要な要素の一つである。ヨットは基本的に風の力から推進力を得るもので、それはセール内外の面を通過する風の気圧差による揚力によって生じる。飛行機の翼の形状のような役割である。従って、風を受けるセールの形状によって同じ風であっても得ることができる推進力は異なってくる。ヨットの大会では指定されたコースをいかに早く周回できるかで優劣が決まるため、ヨットの速さという要素はヨットの選手の間でも優先度の高いものである。著者の経験においても、大学の部活動としてヨット部で活動を行う上で、毎週ヨットの速さを上げる努力を行っていた。

ヨットのセールの形状の確認方法として、大学レベルの活動では実際の海上練習中の様子を動画で撮影し、それを陸に帰ったのちに確認している。動画を用いる理由は、実際にヨットに搭乗している選手からはセールの形状を正確に認識することは困難なためである。基本的にヨットチームは緊急時のレスキュー用に小型船舶を保有しており、練習艇を追従して、日々の練習中のヨットのセールの様子を撮影し、それを元にロープワークなどをどう改善すれば自分のヨットが速くなるかという議論を行っていた。

しかし、動画からセールの形状を正確に把握することはヨット競技を始めたての初心者には難しいものであり、著者の場合には大学1年生からヨットを始めたものの、セールの形状を動画から理解できるようになったのは大学3年生の後半であった。ヨットを始めてからセール形状を動画から理解できるようになるまでの時間は個人差が含まれるもの、その理解ができていない段階では、ほかの競技者によるヨットの速さの向上のために行っている議論のレベルに達せず、上達も遅れることになる。

この問題を解決するために動画の情報からセールの形状を分かりやすく提示することを

考えた。本研究では、ヨットの練習中に撮影された動画の1フレームを対象として、ヨットのセールの中でもメインセールと呼ばれるセールの形状の可視化と、推進力に関するメインセールの、風を受けてはらんでいる内側の面（以後メインセールの内側面と呼ぶ）の可視化を実現することを目的とした。これによってヨット競技の初心者が早くに習熟者の議論のレベルに達し、ヨットのセーリング技術の早期向上が期待される。

1.2 関連研究

ヨットのセールの形状をカメラで撮影された映像から立体形状を復元している研究として、文献[1]がある。この研究では画像内の特徴とするマーカーを数多く直接セールに取り付け、これを単眼カメラで撮影することでセールの形状を復元している。マーカーにはそれぞれに異なる模様が付けられており、その模様の変形の度合いによってセールが風を受けることによる立体形状の推定を行っている。

文献[1]ではセールに大量のマーカーを取り付けており、その結果はセーリングの専門家による定性的評価において、再構築されたセールの立体形状は妥当であると結論づけられた。この手法は比較的安価に再現可能であるが、大学でのヨット競技の活動規模の場合では、セールに何枚ものマーカーを取り付けること自体がセールに小さな損傷を与える可能性があるため、本研究ではセールに変更を加えないこととした。

1.3 本研究の特徴と本論文の概要

本研究では、携帯電話やスマートフォンに内蔵されたカメラによって撮影された動画からヨットのメインセールの形状の可視化を行うシステムの実現を目的とする。対象は、470級（ヨンナナマルキュウ）クラスのヨットのメインセールとし、ヨットが風上に進んでいる動画を後方から撮影した映像を対象とする。撮影した動画の1フレームを入力とし、セールの輪郭や内側面を可視化する。このシステムを利用することで、ヨット初心者のセーリング技術の早期向上を期待できる。

本章に続く各章の概要は次の通りである。

- 第2章では、ヨットの知識や仕組み、本研究で利用した物体検出手法やキーポイント検出手法の仕組み等の基礎知識を示す。
- 第3章では、提案手法の概要と各処理方法について示す。
- 第4章では、メインセールの検出手法とセール内の特徴点に対するキーポイン

ト検出手法の有効性の実験および提案システムを利用して形状の可視化を行った実験とその結果、考察を示す。

- 第5章では、本論文のまとめを行い、今後の課題を示す。

第 2 章

本研究に用いる基本的手法と知識

2.1 ヨットの知識とセールの役割

今回の研究で対象とする 470 級、メインセールの説明とセールの役割を示す。

470 級 [2]

1963 年にデザインされた 2 人乗りのヨットクラスである。名前の由来は全長が 4.7m であることによる。1976 年から五輪種目としても採用されている。

メインセール

今回対象とするメインセールは 470 級が有するセールの中で最も大きいもので、図 2.1 の塗りつぶされている部分である。三角形のような形状をしており、そのうち二辺はマストやブームと繋がっている。繋がっていない辺はリーチと呼ばれる。

セールの役割

ヨットにおけるセールの役割は、進むための力を風から得ることである。文献 [4] に従って説明する。

ヨットは風上を 0° とした場合に 45° 近くの角度まで上ることができる。すなわち、風上方向に進むことができる。ここでは風から 45° の角度で進むヨットの仕組みの説明を行う。図 2.2 のようにヨットに風が当たった場合、ヨットのセールは風を受け膨らみ、飛行機の翼状の形状となり、その結果左右での気流に違いが生じ圧力差が生じる。これにより揚力が発生する。また、揚力とは別に、それを打ち消す力も必要である。それが図 2.1 内にあるセンターボードの部位の働きであり、ヨットが揚力を受けることに対して海水から受ける抵抗力を生じさせる。この抵抗力と揚力が打ち消し合い、図 2.3 のように、進行方向に進むことができる。

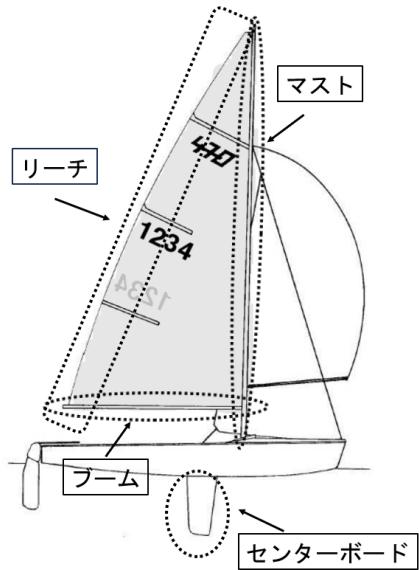


Fig.2.1: 470 級とメインセール (文献 [3] の図を改変)

このように、セールは風から揚力を得ることがヨットにおける役割である。セールは、単に風を受けるだけでなく、気圧差を生じさせ進む力を得る。その力を制御するためには、メインセールの形状は非常に重要である。

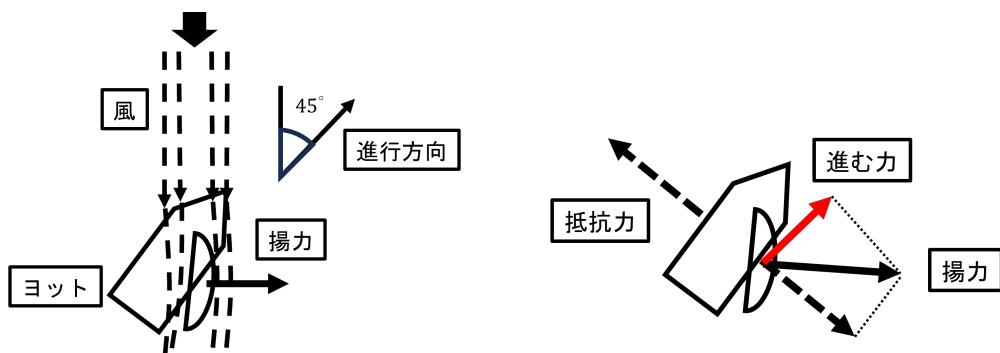


Fig.2.2: 揚力の発生

Fig.2.3: 風上へ進む力の仕組み

なお、ヨットは風上から風下に移動することも可能だが、それは図 2.4 のように、単に風をセールで受け、その力で進むため、セールの形状は風下から風上へ上るときほど重要ではないため、ここでは対象としない。

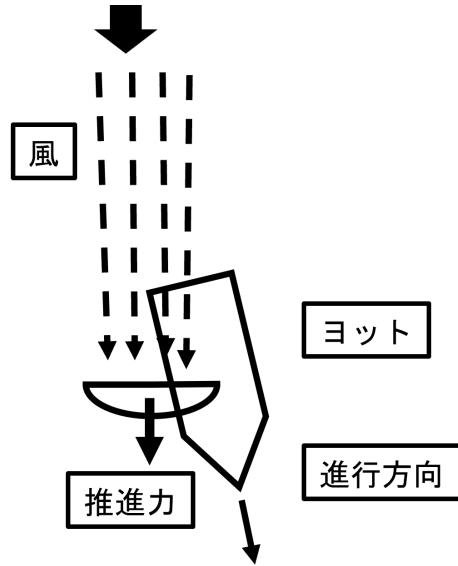


Fig.2.4: 風下へ進む力の仕組み

2.2 メインセールの幾何形状

メインセールは剛体であるマストとブームに 2 端を固定された三角形状の曲面 S であり (図 2.1), 風を受けていない状態ではガウス曲率 0 の平坦な形状をしている。しかし風を受けた状態では風下方向に向かって凸な形状となり, 曲面 S 上のあらゆる点でガウス曲率が正の楕円点となる (図 2.5)。本研究ではメインセールの風上側を「メインセールの内側面」としている。以後メインセールの風下側(凸の向いている方の側)の面を「メインセールの外側面」と呼ぶことにする。

2 次元曲面 S は互いに素な内部点 S^i と境界 ∂S の和集合であるが, 内部点 S^i はさらに画像上で外側面が見えている点 S^{outer} と内側面が見えている点 S^{inner} と画像上では見えていない点 S^{unseen} の互いに素な 3 つの部分集合の和集合となっている。

$$\begin{aligned} S &= S^i \cup \partial S \\ &= S^{\text{outer}} \cup S^{\text{inner}} \cup S^{\text{unseen}} \cup \partial S \end{aligned}$$

メインセール S の外側面/内側面は風の方向に応じて相互に切り替わる。また切り替わりの途中でガウス曲率が 0 に近い状態(遷移状態)を経由すると考えられるが, 本研究では十分に風を受けて S 上の各点でガウス曲率が 0 では無い幾何形状を対象としている。

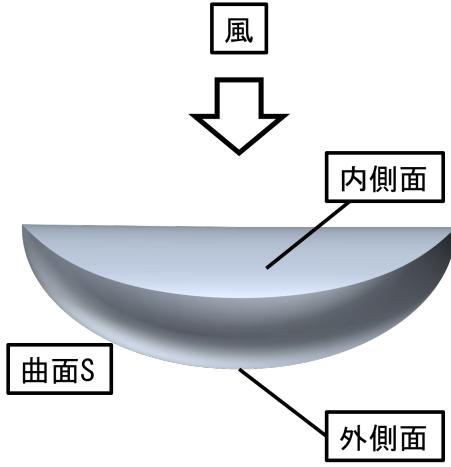


Fig.2.5: 風を受けた曲面 S の形状

この時、剛体であるマストとブームで決定される平面からメインセール S への垂直距離を考え、これを S 上の各点における「メインセールの深さ」と呼ぶ。著者の経験上、メインセールの深さはセール形状を考えるうえで重要な要素であり、風が弱い環境ではセールの外側面を通過する風の剥離が起こりやすく、そのため深さを浅くすることで外側面の風の流れを良くする必要があり、風が強い環境では風の剥離は起きづらいため、深さを深くすることにより大きな揚力を得ることができる。

ヨットの練習動画からメインセールの形状を把握するためには、上記のメインセールの外側面、メインセールの内側面、メインセールの境界、メインセールの深さの最大値とそのセール上での位置を把握することが重要である。しかしメインセールの深さ分布を知るために、3次元形状の復元/3次元計測が必要となり、低解像度のヨット練習動画からは実現困難な課題である。本研究では、画像上で S^{outer} , S^{inner} , ∂S を判別し、マーキングすることでメインセールの幾何形理解を補助し、将来のメインセール深さ計測のための基礎となることを目的としている。

2.3 物体検出手法 (YOLOv5)

物体検出とは、1つの画像の中に何らかの物体が映っている際に、物体の外接矩形(座標、幅、高さ)とその物体のクラスを予測する手法である。同様な概念に物体形状を画素単位で表すマスクを生成するインスタンスセグメンテーション、画像中の全画素のクラスを表すセマンティックセグメンテーションが存在する。

CNN を用いた物体検出手法としては、Faster R-CNN や Mask R-CNN, YOLO など

が挙げられる。R-CNN 系の手法はオブジェクト領域推定を行った後、推定されたそれぞれの候補領域に対してクラス分類が行われるため、検出の精度は良いが、処理が複雑であり、学習にて膨大な時間を要する。一方、YOLO は、上記の 2 つの処理を単一のネットワークで行うことができるため、精度は R-CNN 系と比較すると劣る面もあるが、学習に要する時間は早い。

YOLO とは You only look once (YOLO) : it looks for the frame only once and able to detect all the objects in the frame を意味する。本研究に用いた YOLOv5 は、2020 年 6 月に公開された YOLO モデルの 1 つである。YOLOv5 のネットワーク構造を図 2.6 に示す。大きく分けて Backbone, Neck, Prediction の 3 つのパートに分けられる。

YOLOv5 では、 s , m , l , x の 4 つの学習済みモデルの重みが存在しており、ここでは s を用いた。ここから s の重みを使用した Yolov5s について文献 [5] を用いて説明する。まず、RGB の 3 チャンネルから構成されるカラー画像を Backbone へ入力する。Backbone では、畠み込み演算を行い、個々の画像の特徴マップを作成する。図 2.6 における CBL は、Conv2D, BatchNormal, LeakyReLU を示している。CSP1_x, CSP2_x の内部処理は図 2.7 のような構造である。CSPNet のネットワーク構造を用いて CBL, Res unit, 畠み込み層から構成されていて、 x によって Res unit もしくは CBL の繰り返す数が異なる。そして、文献 [6] によると、SPP (Special Pyramid Pooling) 層で、抽出された特徴量を固定長の特徴ベクトルへ変換している。複数のカーネルサイズで同時にプーリングすることで、詳細かつ広域な情報を取得することが可能である。Neck は、Backbone で生成された特徴マップを組み合わせるレイヤーであり、ここでは Mask R-CNN を元にした PANet と FPNet をベースに改良した bottom-up 型の特徴ピラミッド構造が採用されている。これによって、異なるスケールでの物体を認識することを可能としている。Head では、Neck で生成した特徴マップにより、矩形領域の座標やクラスに関する予測を行う。YOLOv5 では、以上のようなステップを経て物体検出を行う。

YOLOv5 では、学習の際、バッチ毎にデータをデータローダに格納し、データ拡張処理として、その都度スケーリング、色空間の調整、モザイク処理の 3 種類の処理を行う。モザイク処理は、4 つの学習用画像データを特定の比率で結合して 1 つの画像データとする。このことで、モデルは通常よりも小さい縮尺でオブジェクトを識別する方法を学習できる。また、損失関数として、YOLOv5 は *ClassLoss*, *ObjectLoss*, *CIoULoss* を用いている。ここで、*ClassLoss* は予測されたクラスラベルが実際のクラスラベルとどの程度一致しているか、*ObjectLoss* は予測された物体の矩形領域が実際の矩形領域とどの程度一致しているか、*CIoULoss* は予測された矩形領域が実際の矩形領域にどれだけ近いかを測定する。それぞれの計算方法には、*ClassLoss* と *ObjectLoss* では、ともにバイ

ナリクロスエントロピー損失を用いており、式 2.1 のように表すことができる。

$$H(y, \hat{y}) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (2.1)$$

また、*CIoULoss* は CIoU(Complete IoU) Loss を採用しており、式 2.6 で表される。CIoU は 2 つの矩形領域の類似性を評価する指標で、IoU に中心間の距離やアスペクト比の差異を考慮したものである。IoU とは、画像の重なりの割合を表す指標で、式 2.2 で求められる。CIoU の計算には、IoU の他に式 2.3, 2.4, 2.5 を用いる。式 2.3 の v は、予測された矩形領域と実際の物体の矩形領域のアスペクト比の一致度を表す。また、式 2.5 にて用いられている ρ は 2 つの矩形領域の中心間のユークリッド距離であり、 c は凸対角線の長さを表している。*CIoULoss* は、1 から CIoU を引くことで算出される。

$$\text{IoU} = \frac{A \cap B}{A \cup B} \quad (2.2)$$

$$v = \left(\frac{4}{\pi^2} \right) \times \left(\tan^{-1} \left(\frac{w_2}{h_2} \right) - \tan^{-1} \left(\frac{w_1}{h_1} \right) \right)^2 \quad (2.3)$$

$$\alpha = \frac{v}{1 - \text{IoU} + v} \quad (2.4)$$

$$\text{CIoU} = \text{IoU} - \frac{\rho^2}{c^2} - \alpha v \quad (2.5)$$

$$L_{CIoU} = 1 - \text{CIoU} \quad (2.6)$$

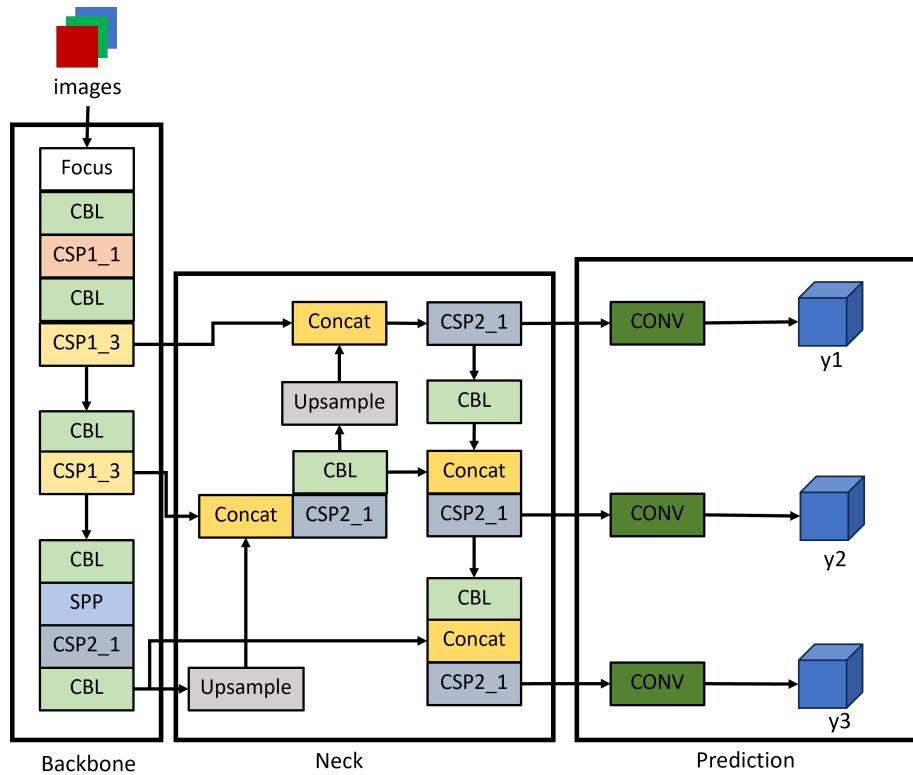


Fig.2.6: YOLOv5 のネットワーク (文献 [5] の図 3 より引用)

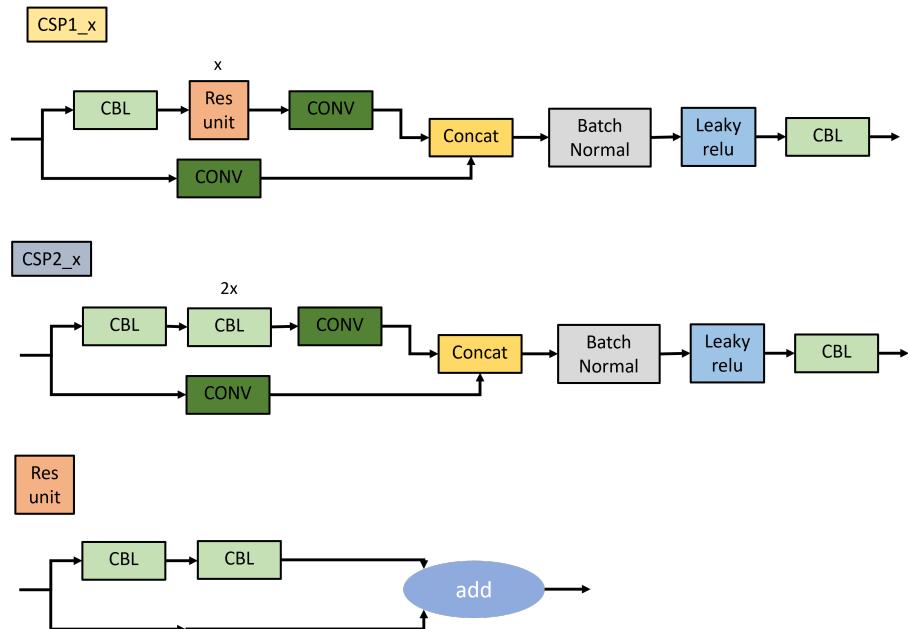


Fig.2.7: CSP のネットワーク (文献 [5] の図 3 より引用)

2.4 キーポイント検出手法 (Mask R-CNN[8])

キーポイント検出とは、画像内に存在する物体の特徴点（キーポイント）を検出手法である。まず、キーポイントは画像内の物体で注目すべき局所的な特徴をあらわす点または座標である。キーポイントは画像内で一意であることが望ましく異なる画像間で対応点として使用され、物体検出、物体追跡、画像認識、パノラマ合成、三次元モデリングなどで広く利用されている。データの形式はキーポイントを持つ物体のある矩形領域（座標、幅、高さ）と、それぞれどのような物体を示しているかというクラスに加えてキーポイントの座標と、それぞれのキーポイントのクラスとする実装がある。

本論文で用いたアルゴリズムとしては detectron2 のライブラリ内に存在する Mask R-CNN を使用している。文献 [7] によると、detectron2 は、最先端の物体検出およびセグメンテーションアルゴリズムを提供する Facebook AI Reserch のライブラリであり、2019 年 11 月に前身の detectron から発展したものである。また 2.3 節にて同様の説明を行っているが、CNN を用いた物体検出手法としては、Faster R-CNN や Mask R-CNN、SSD や YOLO などが挙げられ、Mask R-CNN は同様の検出手法と比較すると、処理が複雑であり膨大な時間を要するが、検出の精度は優れている。Mask R-CNN は 2017 年に発表されたモデルであり、Faster R-CNN を改良したモデルである。基本的な Mask R-CNN は検出対象の物体の座標を矩形領域で出力し、そのクラスラベルを出力することに加えて、検出対象の物体領域をマスク処理を行ってピクセル毎に出力する。Mask R-CNN を用いたキーポイント検出はそのアルゴリズムのマスク処理によるインスタンスセグメンテーションの対象を変更し、1 点のみを検出するようにしたものである。ここで、Mask R-CNN ではキーポイント検出の際にマスク処理によるインスタンスセグメンテーションも同時に学習することもできる。図 2.8 はそれらを同時に学習しているモデルを示した。図 2.8 を元に Mask R-CNN を説明する。

まず、RGB の 3 チャンネルから構成されるカラー画像を Mask R-CNN へ入力する。Mask R-CNN には backbone として FPN (Feature Pyramid Network) があり、畳み込み演算を行い、個々の画像の特徴マップを作成する。FPN は CNN を拡張したもので、CNN では、解像度が低下するに従って精度も低下する傾向があるが、FPN は異なる解像度の入力画像の情報を用いることでその問題に対処した。RPN (Region Proposal Network) は、入力として任意のサイズの画像を受け取り、各々が物体らしさの confidence score を持つ矩形の物体の候補領域を出力する。また RPN は、FPN の特徴マップを出力する畳み込み演算の最後の畳み込み層を共有しており、それによって FPN と同じ特徴

マップ上に候補領域を得ることができる。ROI Align は RPN で出力された候補領域を、バイリニア補間の処理によって同じサイズにするネットワークである。同じサイズに変更するのは、その後マスク処理を行うためである。ここからサイズが変更された候補領域を元に出力を行う。

キーポイント検出ではこの候補領域を CNN と二つの Mask Branch に入力する。CNN からは対象物体の矩形領域とクラスが出力される。また、二つの Mask Branch について述べると、図 2.8 の上部中央の Mask Branch は各候補領域に対してインスタンスセグメンテーションマスクを予測する。Mask Branch は FCN (Fully Convolutional Network) で構成されており、各ピクセルに対してバイナリマスクを生成する。また、このマスクは検出対象のクラスごとに学習されるもので、CNN でのクラス出力によって予測するマスクを選択する。図 2.8 の上部左側の Mask Branch も同様のフレームワークを用いており、各キーポイントを独立した one-hot マスクとして予測する。各ピクセルがキーポイントである確率を予測し、最も高い確率を持つピクセルが 1 となる。Mask R-CNN では以上のようなステップを経てキーポイント検出を行う。

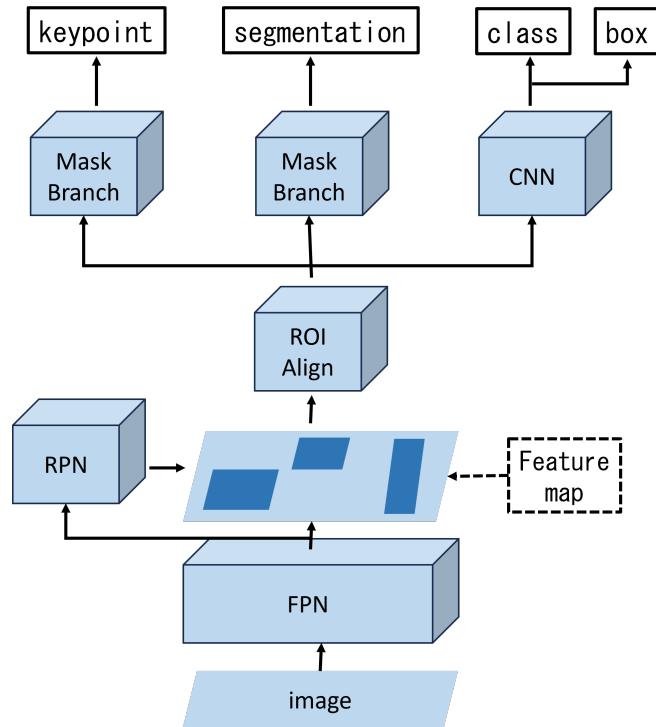


Fig.2.8: Mask R-CNN のネットワーク (文献 [8] の図 1 を改変)

文献 [9] に従って説明する。Mask R-CNN では学習の際に、RPN と Mask R-CNN から RPN を除いたネットワーク (これ以降 “ネットワーク A”と記載) を区別し、4 段階に

分割して学習を行う。その理由として、RPN とネットワーク A はそれぞれ独立であるが、この 2 つは畳み込み層を共有するように学習させるためである。1 段階目では、RPNにおいて、事前に学習されたモデルで初期化し候補領域の提案タスクをエンドツーエンドで学習する。次に 2 段階目として、先ほど学習した RPN が提案した領域を用いて、ネットワーク A の学習を行う。この時、そのネットワークも事前に学習したモデルによって初期化されている。この段階では、まだ RPN とネットワーク A は畳み込み層を共有していない。3 段階目として、RPN を初期化し再度学習させる。この時に先ほど学習したネットワーク A の畳み込み層のうち、RPN と共有するものを固定し、それ以外の層の学習を行う。これにより RPN とネットワーク A は畳み込み層を共有することができる、最後に 4 段階目として、共有した畳み込み層を固定し、ネットワーク A の全結合層をファインチューニングする。

また、学習の際に使用される損失関数もそれぞれ異なる。RPN が考慮する損失は *ObjectnessLoss* と *LocalizationLoss* であり、*ObjectnessLoss* は予測されたオブジェクトの confidence score と実際のラベルとの差異を表し、*LocalizationLoss* は予測された領域と実際の物体領域との差異を表す。*ObjectnessLoss* は式 2.1 に示すバイナリクロスエントロピー損失、*LocalizationLoss* は Smooth L1 損失を用いる。Smooth L1 損失は式 2.7 で表され、ここでの x は予測値から真値を引いた値である。

$$\text{Smooth L1}(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| < 1 \\ |x| - \frac{1}{2} & \text{otherwise} \end{cases} \quad (2.7)$$

一方、ネットワーク A が考慮する損失は、*ClassLoss*, *BoxLoss*, *MaskLoss*, *KeypointLoss* である。ここで、*ClassLoss* は予測されたクラスのラベルと実際のクラスのラベルとの差異、*BoxLoss* は予測された物体の矩形領域と実際の矩形領域との差異、*MaskLoss* は予測された物体領域と実際の物体領域との差異、*KeypointLoss* は予測されたキーポイントと実際にキーポイントとの差異であり、*ClassLoss*, *BoxLoss* は CNN, *MaskLoss*, *KeypointLoss* は Mask Branch で求められる。*ClassLoss* は式 2.8 に示すクロスエントロピー損失によって計算され、*BoxLoss* は式 2.7 に示す Smooth L1 損失を用いる。また、*MaskLoss* は、式 2.1 に示すバイナリクロスエントロピー損失、*KeypointLoss* は式 2.8 に示すクロスエントロピー損失を用いる。

$$H(y, \hat{y}) = - \sum_i y_i \log(\hat{y}_i) \quad (2.8)$$

2.5 検定方法

本研究では、Mask R-CNN のキーポイント検出モデルの比較のために検定を行った。検定には F 検定と t 検定を使用した。

F 検定についての説明を行う。両側検定の帰無仮説と対立仮説は、以下のようなになる。

帰無仮説 2つの分布の母分散は等しい

対立仮説 2つの分布の母分散は等しくない

大きさ N_1 と N_2 の 2 つの標本 $\{x_{11}, x_{12}, \dots, x_{1N_1}\}$, $\{x_{21}, x_{22}, \dots, x_{2N_2}\}$ による母分散の差の検定統計量 $T(s_1^2, s_2^2)$ は

$$T(s_1^2, s_2^2) = \frac{s_1^2}{s_2^2} \quad (2.9)$$

で求めることができる。ただし、

$$s_1^2 = \frac{\sum_{i=1}^{N_1} (x_{1i} - \bar{x}_1)^2}{N_1 - 1} \quad (2.10)$$

$$s_2^2 = \frac{\sum_{i=1}^{N_2} (x_{2i} - \bar{x}_2)^2}{N_2 - 1} \quad (2.11)$$

である。 \bar{x}_1 , \bar{x}_2 は 2 つの標本の平均値である。この時 $T(s_1^2, s_2^2)$ は自由度 $(N_1 - 1, N_2 - 1)$ の F 分布に従う。F 検定にて 2 つの分布の分散が等しいかの検定を行い、その検定結果によって 2 つの分布の平均値の検定を行った。

2 つの分布の分散が等しいと言えるとき、2 つの分布の平均の検定には t 検定を用いた。両側検定の帰無仮説と対立仮説は、以下のようなになる。

帰無仮説 2つの分布の母平均は等しい

対立仮説 2つの分布の母平均は等しくない

大きさ N_1 と N_2 の 2 つの標本 $\{x_{11}, x_{12}, \dots, x_{1N_1}\}$, $\{x_{21}, x_{22}, \dots, x_{2N_2}\}$ による母平均の差 $\mu_1 - \mu_2$ の検定統計量 $T(\bar{x}_1, \bar{x}_2, s^2)$ は

$$T(\bar{x}_1, \bar{x}_2, s^2) = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\frac{1}{N_1} + \frac{1}{N_2})s^2}} \quad (2.12)$$

である。ただし

$$s^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \quad (2.13)$$

で、 $T(\bar{x}_1, \bar{x}_2, s^2)$ の分布は自由度 $N_1 + N_2 - 2$ の t 分布に従う。ここでの s_1, s_2 は式 2.10 と 2.11 である。

2 つの分布の分散が等しいと言えないとき、2 グループの平均の検定には t 検定の中でもウェルチの検定を用いる。両側検定の帰無仮説と対立仮説は、以下のようになる。

帰無仮説 2 つの分布の母平均は等しい

対立仮説 2 つの分布の母平均は等しくない

大きさ N_1 と N_2 の 2 つの標本 $\{x_{11}, x_{12}, \dots, x_{1N_1}\}$, $\{x_{21}, x_{22}, \dots, x_{2N_2}\}$ による母平均の差 $\mu_1 - \mu_2$ の検定統計量 $T(\bar{x}_1, \bar{x}_2, s_1^2, s_2^2)$ は

$$T(\bar{x}_1, \bar{x}_2, s_1^2, s_2^2) = \frac{\bar{x}_1 - \bar{x}_2}{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)} \quad (2.14)$$

である。また、ここでの s_1, s_2 は式 2.10 と 2.11 である。これは自由度 m の t 分布に従う。この自由度 m は

$$m = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\sqrt{\frac{s_1^4}{N_1^2(N_1-1)} + \frac{s_2^4}{N_2^2(N_2-1)}}} \quad (2.15)$$

で、 $T(\bar{x}_1, \bar{x}_2, s^2)$ の分布は自由度 $N_1 + N_2$ の t 分布に従う。

m が整数でないとき。最も近い整数値を m とする。

検定は全て有意水準 5% で行った。

第3章

メインセールの形状の可視化システム

3.1 提案システムの概要

本システムは、2.2節で述べたように、メインセール輪郭の可視化とメインセールの外側面と内側面の区別のために内側面のみの着色を行うことを目的としている。そのために入力された画像データからメインセールの物体検出、およびメインセールのキーポイント検出を行う。ヨットに乗る際に、ヨットの推進力にはセールが大きくかかわっていることは第2章にて説明したが、ヨット競技を始めたばかりの選手にとって、セールの画像を見てもそのセールの形状の理解が困難である。よって、このシステムではメインセールが存在する画像に、メインセールの輪郭を可視化させることで、経験の浅い選手がセールの形状を把握する支援を行う。また、メインセールの内側面に着色を行うことで、メインセール形状の把握がより進展することを期待する。

本研究で提案するシステムの流れについて、図3.1を元に説明を行う。まず、与えられた入力画像に対して、YOLOv5にてメインセールの物体検出を行い、検出されたメインセールを囲む矩形領域を抽出する。これは、ヨットが映っている部分を拡大した画像を得るためにあり、ヨットが大きく映ったデータと小さく映ったデータのいずれにも対応できるための処理である。次に抽出した画像に対し、detectron2ライブラリ内のMask R-CNNにてキーポイント検出を行う。ここで検出するキーポイントは図3.2中に示された赤点の8点であり、全てメインセールの輪郭上に存在する。以後、これらのキーポイントを図3.2に示されたラベルで表す。また、メインセールの内側面について着色を行う。

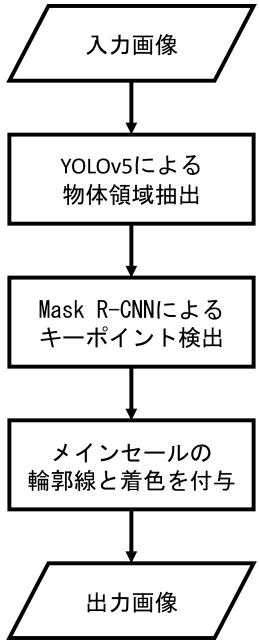


Fig.3.1: 提案システムの流れ

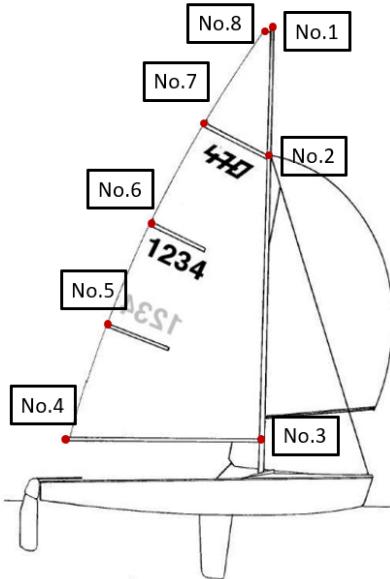


Fig.3.2: 検出するキーポイント (文献 [3] の図を改変)

3.2 YOLOv5 による物体検出

与えられた入力画像に対して、メインセールを検出し、その物体領域を矩形領域で抽出する。YOLOv5 は入力として画像を与えると、学習した複数の物体に対して画像中で検出した物体の confidence score とその物体領域の外接矩形がその中心の x, y 座標と幅、高さの 4 つのパラメータで出力される。検出された物体領域の中で最も高い confidence score を持つ矩形領域を抽出する。物体検出に使用する初期重みは、YOLOv5 のライブラリセットに含まれる学習済みの重みを用いた。ここで扱う学習済みの重みは、ウェブサイト [11] にて公開されている “COCO 2017” にて学習した “Yolov5s” を使用し、それをヨット画像にてファインチューニングする。

3.3 Mask R-CNN によるキーポイント検出

与えられた入力画像に対して、3.2 節の処理でメインセールを検出を行った。メインセールの画像について、ここではその物体領域を矩形領域で検出し、その矩形領域内でキーポイントを検出する。Mask R-CNN は入力として画像を与えると、検出した物体のクラスラベルと confidence score、その物体領域の矩形領域、キーポイントの座標を出力

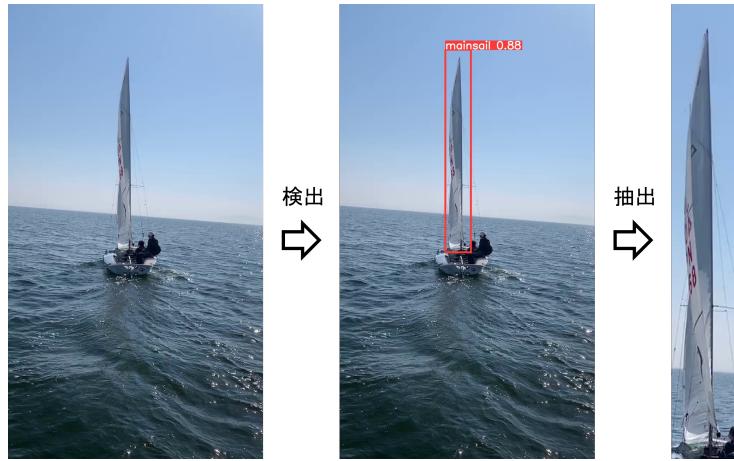


Fig.3.3: YOLOv5 によるメインセールの抽出

する。任意でインスタンスセグメンテーションの学習も可能である。その中のキーポイントの座標を利用する。複数の物体の矩形領域が検出された場合は、最も confidence score の高い矩形領域内で検出されたキーポイントの座標を取得する。物体検出に使用する初期の重みは、detectron2 ライブラリセットに含まれる学習済みの重みとする。ここで扱う学習済みの重みは、初期重みとして論文 [10] にて公開されている ResNet-50 をウェブサイト [11] にて公開されている “COCO 2017” にて最大 12epoch 学習したモデルであり、それをヨット画像にてファインチューニングする。

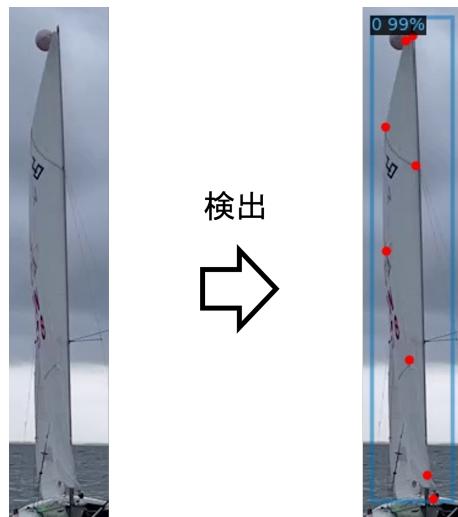


Fig.3.4: Mask R-CNN によるメインセールのキーポイント検出

3.4 メインセールの形状可視化システム

3.3 節の処理によって検出されたキーポイントより、メインセールの形状を可視化する。本システムにて入力画像にメインセールの輪郭と内側面の着色を付与する。

3.4.1 輪郭線の推定アルゴリズム

輪郭可視化システムを図 3.5 を元に説明する。3.3 節の処理にて得たキーポイント座標を用いて、輪郭の可視化を行う。与えられたキーポイントの検出座標を元に、図 3.2 にて示したキーポイント番号の差が 1 であるキーポイント同士を結ぶ線分により、輪郭の可視化を行う。また、キーポイント番号が 1 と 8 を線分として結ぶ。

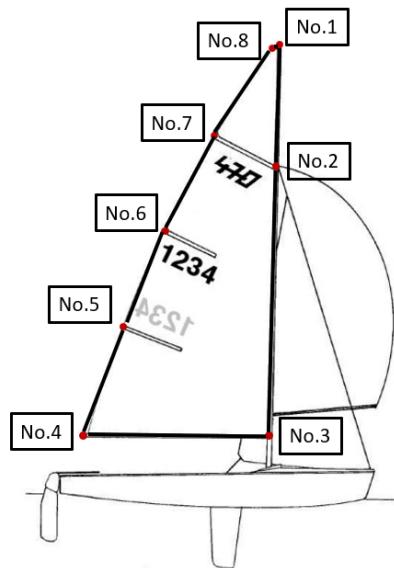


Fig.3.5: 輪郭の可視化のイメージ (文献 [3] の図を改変)

3.4.2 内側面の推定アルゴリズム

メインセールの内側面を着色する。図 3.2 に示したキーポイント番号 1 と 2, もしくは 2 と 3 を結ぶ線分のいずれかがキーポイント番号 4 と 5, 5 と 6, 6 と 7 を結ぶ線分のいずれかと交点を持つとき、その交点と、交点よりも y 座標が小さいキーポイントに囲まれた

領域を内側面とする。また、交点がない場合は8つのキーポイントに囲まれた領域を内側面とする。ここでは、座標は画像の左上の点を原点(0,0)とし、右方向をx座標、また下方向をy座標と設定している。この処理は、以下の仮定に基づく。

仮定1 メインセールの内側面が映っていない画像は入力されない

仮定2 キーポイント番号8と7を結ぶ線分はキーポイント番号1と2、2と3を結ぶ線分とは交わらない

本実験にて扱う動画について、ヨットの中心軸の延長線上にカメラが存在する状態でメインセールにて風をはらむため、メインセール上部は仮定2を満たすと考えられる。

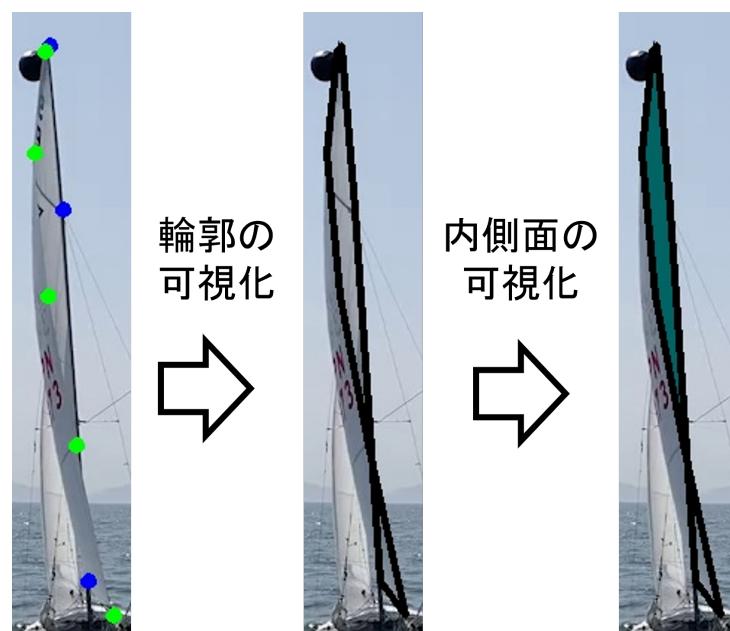


Fig.3.6: 形状可視化システム

第 4 章

実験

4.1 YOLOv5 によるメインセール領域の検出

4.1.1 方法

ここでは、YOLOv5 による物体検出によりメインセールの検出性能を評価する。実際の練習動画により YOLOv5 を学習させ、その精度を求める実験を行った。また、本実験では、全ての画像において映っているヨットは 1 隻のみである。

データセットについて、本研究では名古屋大学体育会ヨット部にて 2020 年から 2024 年までに撮影された練習動画を使用した。動画の中から、ヨットが大きく撮影された 35 編の動画を選択し、そのうちの 22 編の動画から機械的に 10 秒に 1 枚間隔で画像を抽出し、残りの 13 編の動画からは 1 フレーム目を抽出した。その結果、22 編の動画からは 145 枚の画像、13 編の動画からは 13 枚の画像を得た。22 編の動画から得た 145 枚のうち、114 枚を学習用データセット、31 枚を検証用データセットとし、13 編の動画から得た 13 枚をテスト用データセットとした。

学習用データセット、検証用データセットに対し手動にて処理を行った。アノテーション処理にはアノテーションツールである LabelImg [12] を使用した。LabelImg によるアノテーションの表示例を図 4.1 を示す。LabelImg は、各画像において、ラベル付けを行う物体を囲むように手動で矩形領域を指定しラベル付けを行う。このアノテーション処理によって、画像とテキストファイル(クラスラベルと矩形領域の中心座標、矩形領域の高さと幅)のデータセットが得られる。

テスト用データセットに対しても手動にてアノテーション処理を行った。ここで、さらにデータ拡張によって、サンプル数を 13 枚から 78 枚に増加させた。使用したデータ拡張

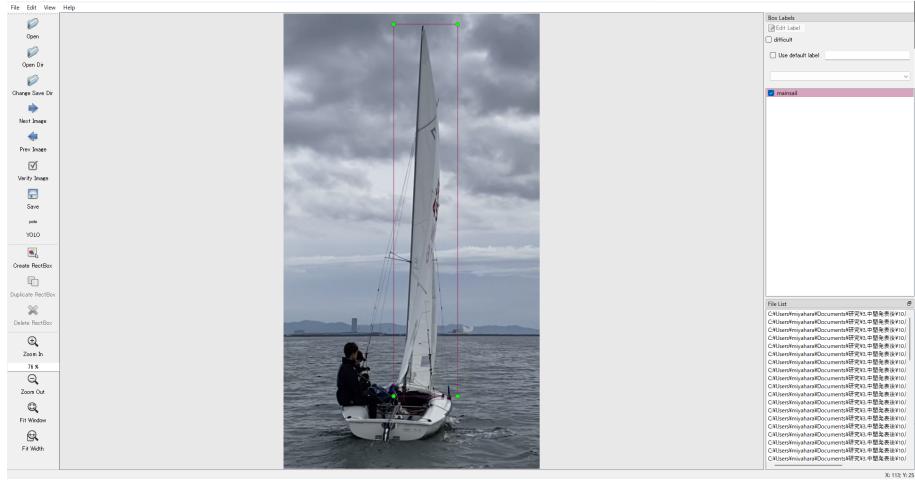


Fig.4.1: labelImg によるアノテーション処理での表示例

手法については、

- 水平方向に反転させる
- 時計回り、または反時計回りに 5° 回転させる

を行った。まず初めに、全てのテストデータ 13 枚を水平方向に反転させた。元のものと、反転したものについて時計回りと反時計回りにそれぞれ 5° 回転させることで元のものと合わせて 78 枚とした。アノテーション処理には、coco-annotator [13] を使用した。coco-annotator では多角形の頂点の座標をアノテーションすることでその外接矩形の左上の座標と幅、高さを得ることができるため、coco-annotator にてメインセールの外接矩形を多角形の頂点をアノテーションすることで得た。coco-annotator によるアノテーションの処理の表示例を図 4.2 に示す。

114 枚の学習用データセットと 31 枚の検証用データセットを使用し、YOLOv5 の学習を行った。学習時の条件は、epoch 数を 300、batch-size を 8 とし、各画像に対してクラス数は 1 を設定した。また、YOLOv5 から提供されている学習済み重み (YOLOv5s) からファインチューニングを行った。ここで、YOLOv5 の出力はクラスラベルと confidence score、物体の矩形領域であり、1 つの入力画像に対する物体検出の出力は最も confidence score の高いもののクラスラベルと矩形領域とした。これは、全ての画像内にヨットは 1 隻映っているとの仮定に基づき、尤もらしい矩形領域をメインセール抽出に使用するためである。また、学習中に検証用データセットを用いて、検出された矩形領域と真値との IoU を計算し、その平均精度が最も高いモデルを結果として得た。

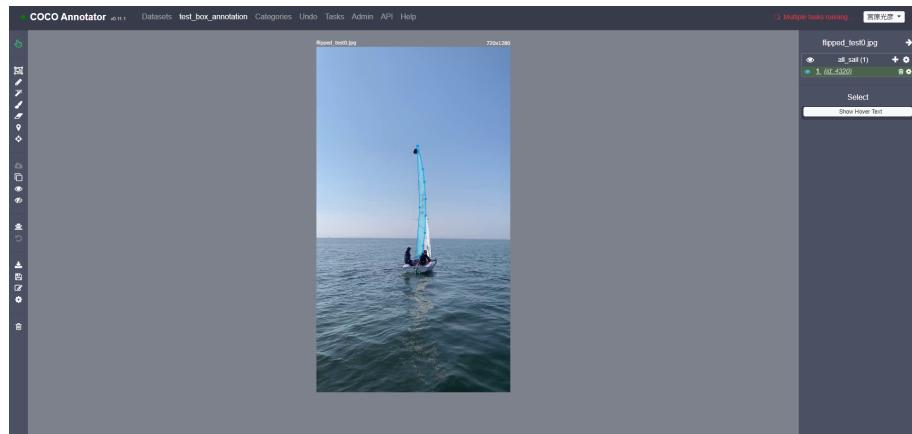


Fig.4.2: coco-annotator によるアノテーション処理での表示例

4.1.2 結果

実際に学習を行った YOLOv5 の出力例は図 4.3 のようになった.

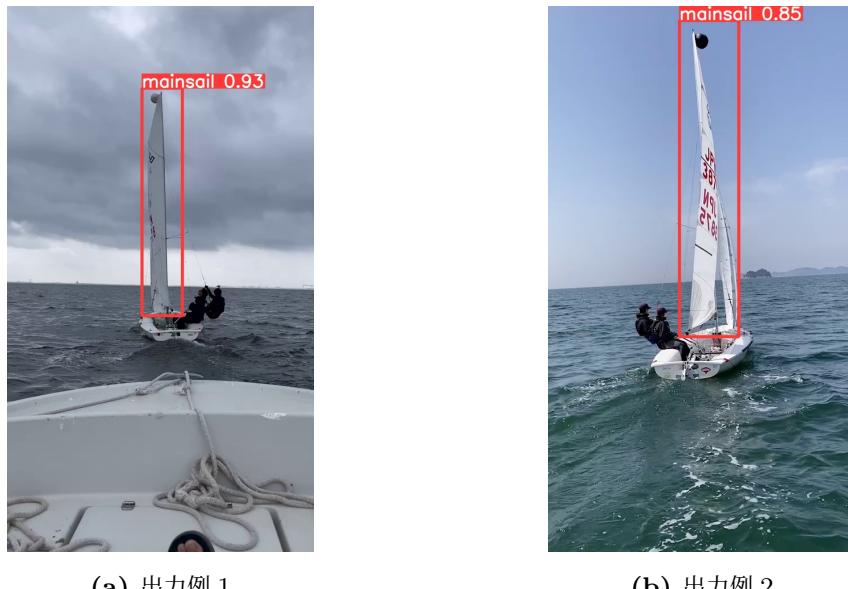


Fig.4.3: YOLOv5 の出力例

学習によって得られたモデルについての Recall による評価を行った. 検出処理において、1つの入力画像に対して、その出力は最も confidence score の高いものとしたため、出力数は1画像につき1以下である. このことより、Recall での評価とした. 検出の正

しさについては、テストデータセットのメインセールの物体領域が検出した矩形領域内にすべて含まれているかを確認し、すべて含まれていればその検出を正しいと定義した。ここで、メインセールの物体領域は coco-annotator によるアノテーション処理にて得たものとする。

セール領域を含むかの判定と精度は表 4.1 のようである。評価の結果、78 セットのテストデータセットのうち、検出結果が正しいとされたものは 71 枚、正しくないとされたものは 7 枚であった。これにより、recall の値は 0.91 となった。

Table 4.1: YOLOv5 によるセール領域の検出結果と検出精度

判定	数値
真	71
偽	7
精度	数値
Recall	0.91

4.1.3 考察

今回は、テストデータセットに反転と $\pm 5^\circ$ 回転のデータ拡張処理を行った。それにより、より一般のテストデータセットによる評価を行うことで、recall の値もより正確になることを期待した。その一方で今回 False と判別されたテストデータは全てそのデータ拡張手法によって生成されたものであった。学習用データセット、検証用データセットにはデータ拡張処理を適用しておらず、今回のモデルでは検出できないテストデータを使用していた可能性がある。

また、本実験によって検出されたメインセールの矩形領域を視覚的に確認した結果、矩形領域がメインセールをすべて含んでいるが False と判別されたものが存在した。検出結果の True もしくは False の判別に使用している正解ラベルは手動で行ったアノテーション処理によるものであるため、真値としたメインセールの物体領域と実際の物体領域にはある程度の誤差が生じている可能性がある。また、True と判別された検出結果に関しては、実際に矩形領域がメインセールの物体領域の一部を含んでいない結果は視覚的には確認できなかった。

4.2 Mask R-CNN によるキーポイントの検出

4.2.1 実験方法

Mask R-CNN によるメインセール内のキーポイント検出精度に対する評価実験を行った。本実験において、全ての画像に対し、映っているヨットは 1 隻とする。

データセット

データセットについて、本研究ではメインセール領域の抽出実験と同様に名古屋大学体育会ヨット部にて 2020 年から 2024 年までに撮影された練習動画を使用した。用意した動画の中から、ヨットが大きく撮影された 36 編の動画を選択し、そのうち 23 編の動画から 10 秒に 1 枚の間隔で画像を抽出し、残りの 13 編の動画からは 1 フレーム目を抽出した。その結果、23 編の動画からは 148 枚の画像、13 編の動画からは 13 枚の画像を抽出した。この 148 枚の画像を学習用データセットとし、13 枚の画像をテスト用データセットとした。テスト用データセットについて 4.1 節と同様のデータ拡張処理を行い、78 枚とした。その後、学習用データセット、テスト用データセットに対し 4.1 節に述べた実験による YOLOv5 の学習済みモデルによるメインセールの物体検出を行い、そのセール領域を抽出した。その際にメインセールの検出が 4.1 節の実験で False とされたものは、その画像サイズの高さの 50 分の 1 の画素数分上下左右に拡大した矩形領域で抽出することで、データセットの画像全てでメインセールが大きく映り、かつメインセールがすべて含まれる画像とすることができた。

学習用データセットに対し、アノテーション処理を行った。アノテーション処理にはアノテーションツールである coco-annotator を使用した。coco-annotator によるアノテーションの処理の表示例を図 4.5 に示す。coco-annotator は、各画像において、ラベル付けを行いたい物体に対して手動で物体領域を特定し、ラベル付けを行う。その後、その物体領域内にキーポイントの座標を手動で特定する。本実験では学習用データセットに対し複数のアノテーションを行い、複数の学習用データセットを作成した。ここで、物体検出でのクラスラベルは図 4.4a にて示すメインセール、マスト、ブーム、リーチとした。アノテーション処理にて与えるのは物体の外接矩形とビットマスク、キーポイント座標であり、外接矩形は物体の矩形領域、ビットマスクはインスタンスセグメンテーション、キーポイント座標はキーポイント検出の学習に用いる。coco-annotator では多角形の頂点の座標をアノテーションすることで、自動的に多角形の外接矩形とインスタンスセグメン

テーション用のビットマスクを生成するため、手動にてメインセールの物体領域を含む多角形の頂点とキーポイント座標をアノテーションした。このアノテーション処理により、148枚の学習用データセットを3セット用意した。ここで、それぞれのクラスが持つキーポイントは表4.2のようになる。キーポイントラベルは図4.4bに従う。

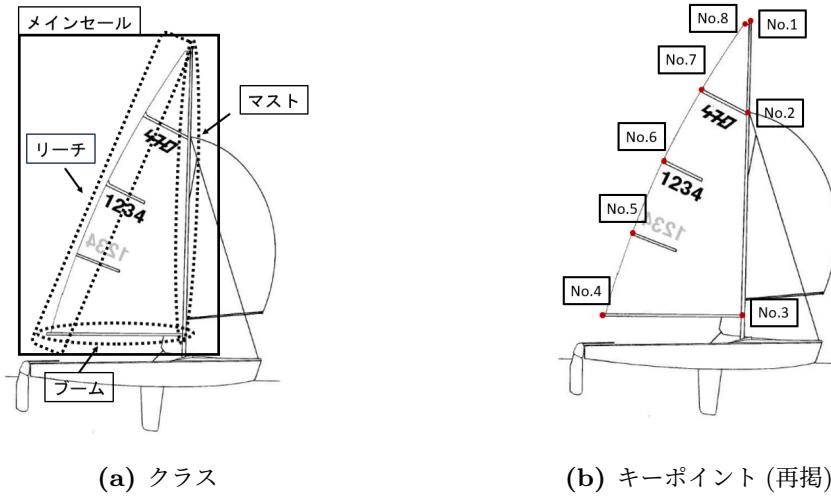


Fig.4.4: 検出するクラスとキーポイント (文献 [3] の図を改変)

Table 4.2: クラスが持つキーポイント

クラスラベル	保持するキーポイント (No)
メインセール	1, 2, 3, 4, 5, 6, 7, 8
マスト	1, 2, 3
ブーム	3, 4
リーチ	4, 5, 6, 7, 8

以下に学習用データセットを示す。

Train1

メインセール1クラスの外接矩形とビットマスク、キーポイント座標をアノテーション処理したもの

Train2

メインセールの境界であるマスト、ブーム、リーチの3種類のみからなるデータセットであり、以下を含む。

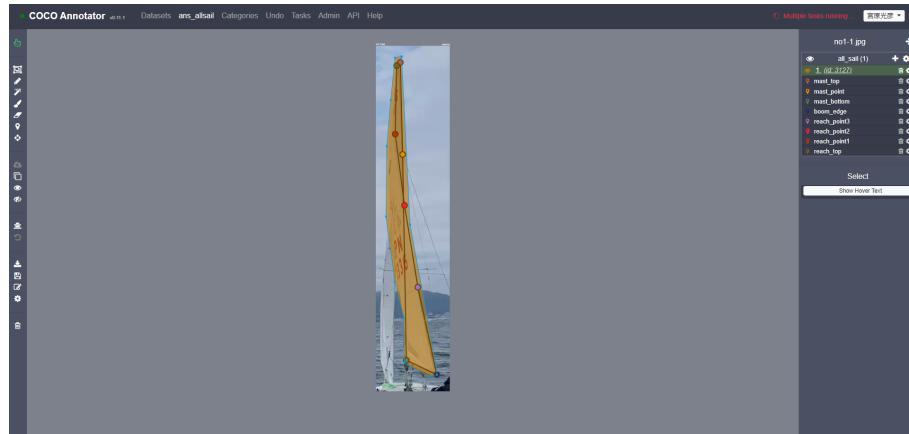


Fig.4.5: coco-annotator によるアノテーション処理での表示例

- マスト 1 クラスの外接矩形とビットマスク, キーポイント座標をアノテーション処理したもの
- ブーム 1 クラスの外接矩形とビットマスク, キーポイント座標をアノテーション処理したもの
- リーチ 1 クラスの外接矩形とビットマスク, キーポイント座標をアノテーション処理したもの

Train1a

Train1 にデータ拡張を行ったもの

Train1 と Train2 は画像データ, キーポイント座標は同じであるが, アノテーションの外接矩形, インスタンスセグメンテーションマスクとクラス数が異なっている. Train1, Train1a のアノテーション例を図 4.6a, Train2 を図 4.6b に示す.

ここで, Train1a にて行ったデータ拡張処理について, 本実験でデータ拡張処理を行うために使用した関数を以下に示す. この関数は detectron2 が定義しているライブラリ関数であり, その関数が学習用データセットが 1 枚呼び出されるごとに適用される.

RandomFlip (どのモデルでも適用される)

画像をランダムに垂直方向もしくは水平方向に反転させる. 0.5 の確率で水平方向に反転させることのみ適用.

ResizeShortestEdge

指定した数字の中からランダムで一つの値が選択され, 画像の最短辺の長さをその数字に合わせる. 扱っている学習画像は 23 編の動画から合計 148 枚の画像を抽出



(a) Train1, 1a



(b) Train2

Fig.4.6: それぞれのアノテーション例

していることから、この 148 枚の画像を抽出元の動画が同じものでグループ分けし、23 組のグループを作成した。そして、その 23 グループごとの最短辺の平均値の平均は 124.01 であった。この平均値を四捨五入し整数とした 124 を基準として、[99, 109, 119, 129, 139, 149] の 6 つの数を指定し、この中からランダムに学習画像の最短辺を合わせる長さを選択する。学習結果が画素数が小さすぎるものや大きすぎるものに偏らないようにするために適用した。

RandomRotation

画像をランダムに回転させる。回転させる角度は学習画像を -5° から 5° の間でランダムに適用させている。ヨットは海上において高頻度で傾きが変動することから適用した。

RandomLighting

画像にランダムな照明の変化を適用する。ヨットの動画は屋外で撮影されるため照明の条件は常に一定ではないことを考慮して適用した。

ネットワーク構成

モデルの学習を行うネットワークアーキテクチャも複数用意した。学習時の条件は、epoch 数を 30、batch-size を 2 とした。また、学習率は 0.01 とし、学習時に 1 枚の画像

に対し, 学習に考慮する候補領域数は 512 とし, detectron2 から提供されている学習済み重みからファインチューニングを行った. 出力の閾値を confidence score が 0.5 より大きいとした. 本実験では以後の処理でキーポイントのみを利用するため, 物体領域のインスタンスセグメンテーションは必ずしも必要ではない. また, キーポイントはセール境界上にのみ存在するため, セール内部点の抽出も必ずしも必要とはいえない. そこでインスタンスセグメンテーションの有無やセール境界のマスト, ブーム, リーチの分類についてクラス数を変化させた複数のアーキテクチャを比較する. 以下にネットワークアーキテクチャを示す.

ネットワーク C1

1 クラスの学習を行う. その際に, 物体検出とキーポイント検出の学習を行う.

ネットワーク ISC1

1 クラスの学習を行う. その際に, 物体検出とインスタンスセグメンテーション, キーポイント検出の学習を行う.

ネットワーク ISC3

3 クラスの学習を行う. その際に, 物体検出とインスタンスセグメンテーション, キーポイント検出の学習を行う.

ここで, ネットワーク 3 について, detectron2 が提供する Mask R-CNN では同一モデルで複数のクラス検出を行うモデルについて, キーポイント検出の学習も行う場合, それらのクラスが持つキーポイントは同一である必要がある. このことから, マスト, ブーム, リーチはそれぞれ 1, 2, 3, 4, 5 というキーポイントラベルのキーポイントを持つとし, 表 4.3 のように割り当てた. マストクラス, ブームクラスを割り当てられないキーポイントラベルについては座標 (0, 0) のアノテーションを与えた.

Table 4.3: キーポイントラベルの割り振り

キーポイントラベル	マスト	ブーム	リーチ
1	1	3	8
2	2		7
3			6
4			5
5	3	4	4

これらの複数の学習用データセットとネットワークアーキテクチャを用いて, 複数のモ

モデルを学習した。

評価方法

学習済みモデルのキーポイント検出結果の比較には、2.5節で述べた有意水準5%でのF検定、t検定や検出数を用いた。検定では、8点のキーポイントごとに、それぞれのキーポイントでの、検出座標と真値の座標とのx座標とy座標それぞれの差の平均値の比較を行った。検出座標は最もconfidence scoreの高かった物体検出におけるキーポイントの検出座標を使用した。ここで、今回のテストデータセットはそれぞれ異なるサイズの画像データより、その正規化を行う必要があった。ここでは、図4.4bで示したキーポイント1と3の座標の真値のユークリッド距離(マスト部分の長さ)を求め、その距離を基準距離 ℓ とし式4.1、4.2のように算出した。

$$x\text{座標の差} = \frac{500}{\text{距離}\ell} \times (\text{検出した座標の}x\text{座標} - \text{真値の座標の}x\text{座標}) \quad (4.1)$$

$$y\text{座標の差} = \frac{500}{\text{距離}\ell} \times (\text{検出した座標の}y\text{座標} - \text{真値の座標の}y\text{座標}) \quad (4.2)$$

検定を行うにあたり、扱う差の分布について、検出されるキーポイントの座標は、真値のキーポイントの座標を中心に分布し、真値のキーポイントの座標付近に分布し、正規性を持つと仮定した。

2つのモデルの比較について、F検定の検定対象には検出座標と真値の座標とのx座標とy座標それぞれの差の分布、t検定には検定対象には検出座標と真値の座標とのx座標とy座標それぞれの差の分布の平均値とした。F検定にて2つの分布の等分散性の検定を行い、その結果を元にt検定の手法を決定した。t検定にて有意差が見られた場合は、平均値が0に近いモデルが有意水準5%で優れている可能性を持つとした。本実験にて行うのは以下の実験である。

実験1 インスタンスセグメンテーション学習の有りと無しの比較

実験2 マスト、ブーム、リーチを1ネットワークで学習するか3ネットワークで学習するかの比較

実験3 物体領域についてメインセールとマスト、ブーム、リーチとの比較

実験4 データ拡張有りと無しの比較

4.2.2 実験 1：インスタンスセグメンテーション学習の有りと無しの比較

本実験では、Mask R-CNN を学習する際のモデルのタスクとして、インスタンスセグメンテーションを学習することによる精度の変化を比較した。

表 4.4 に、He らの研究での実験結果を示す。He ら [8] によると、Mask R-CNN にて、インスタンスセグメンテーションの学習を行ったモデルと行わなかったモデルでは、インスタンスセグメンテーション学習を行ったモデルの方がキーポイント検出の精度がよいとされている。ここで、精度として用いられている AP は平均適合率であり、PR 曲線と縦軸、横軸との面積の評価値である。PR 曲線とは、モデル評価指標である Precision を縦軸、Recall を横軸とした曲線である。 AP_{person}^{bb} は人間検出モデルにおける物体検出の AP を、 AP_{person}^{mask} は人間検出モデルにおけるインスタンスセグメンテーションの AP を、 AP^{kp} は人間検出モデルにおけるキーポイント検出の AP を表している。値は百分率である。

Table 4.4: Mask R-CNN の学習精度 [8]

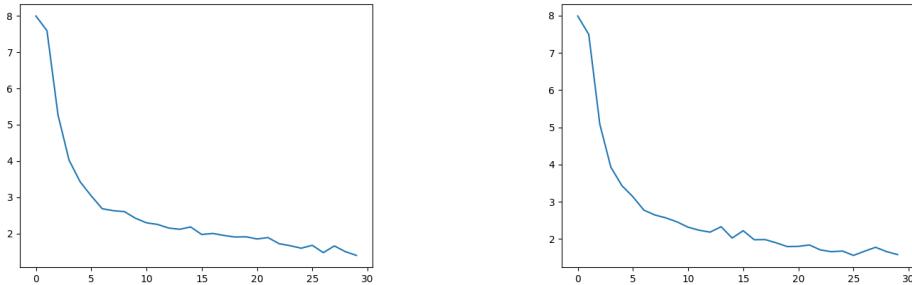
	AP_{person}^{bb}	AP_{person}^{mask}	AP^{kp}
mask 学習無	50.7	-	64.2
mask 学習有	52.0	45.1	64.7

本実験では、インスタンスセグメンテーションを学習することによるキーポイント検出の精度向上が、本研究にもみられるかを確かめることを目的とする。

比較するモデル

本実験にて比較するモデルは、4.2.1 節にて述べたインスタンスセグメンテーションを行わないネットワーク C1 (モデル 1) と行うネットワーク ISC1 (モデル 2) であり、訓練データには Train1 を利用する。

以下の図 4.7 は 2 つのモデルの学習におけるキーポイント検出器の損失関数の値の推移である。学習途中で検証を行っていないため、学習途中での学習不足や過学習の可能性は否めないが、図 4.7 にて、損失関数の値の変化が少ない領域に達していると言える。

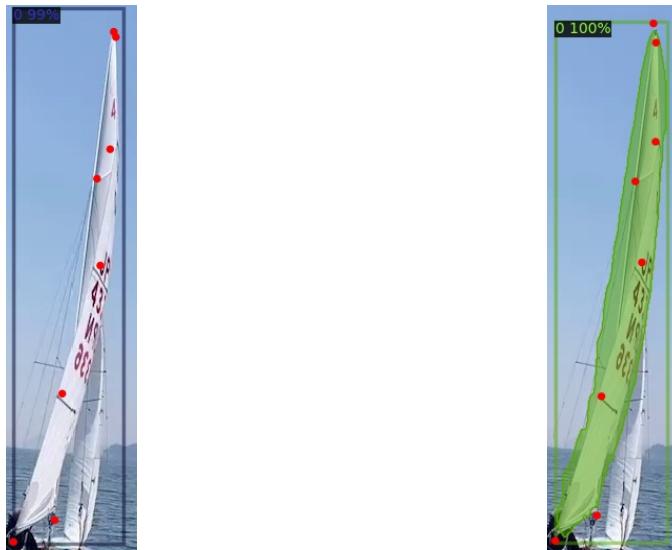


(a) インスタンスセグメンテーションの学習を行
わない場合 (b) インスタンスセグメンテーションの学習を行
う場合

Fig.4.7: 損失関数の値の推移

結果

ネットワーク C1 による出力例は図 4.8a, ネットワーク ISC1 による出力例は図 4.8b のようになった.



(a) インスタンスセグメンテーションの学習を行
わないモデル (b) インスタンスセグメンテーションの学習を行
うモデル

Fig.4.8: 実際の出力例

2 つのモデルが持つキーポイント検出結果を用いて検定を行った. ここでは, 2 つのモデルの検出結果と真値との差の分布は正規性を持つと仮定している. モデル 1 にてキーポ

イント検出を行った際の x 軸, y 軸における真値との差を正規化した数値の分布は図 4.9 の様になり, モデル 2 では図 4.10 のようになった. 図より, これらの分布は正規分布の特性を示唆していることが確認できた.

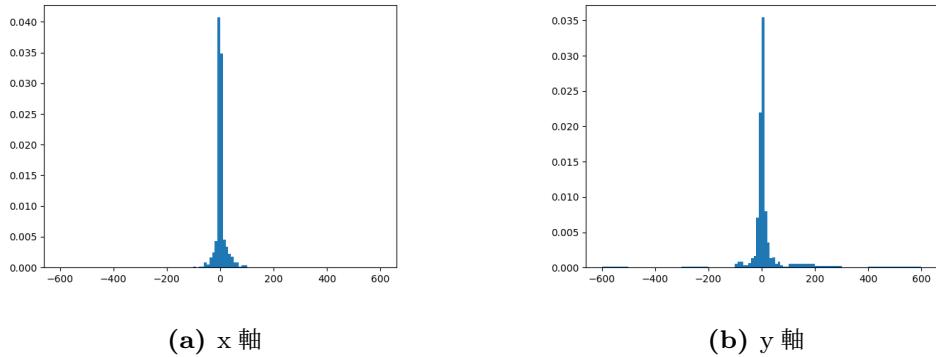


Fig.4.9: インスタンスセグメンテーションを学習しなかったモデルの真値との差の分布

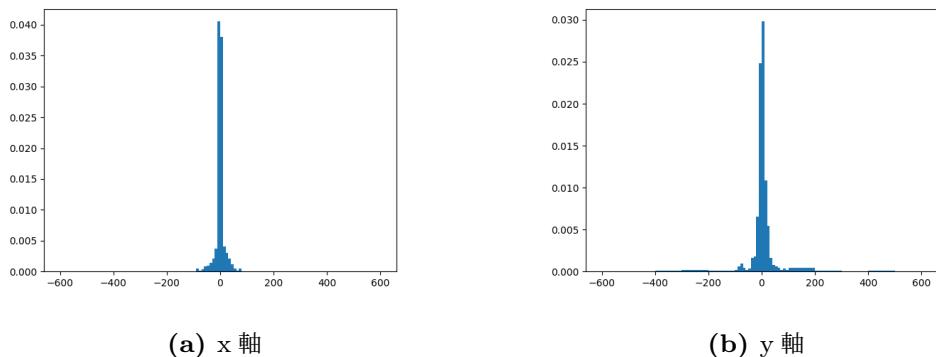


Fig.4.10: インスタンスセグメンテーションを学習したモデルの真値との差の分布

検出数について, どちらのモデルも 78 枚のテストデータセットのすべてにおいて閾値より大きい confidence score を持つ出力が確認された.

F 検定の結果を表 4.5 に示す. この結果を元に t 検定を行った.

t 検定の結果を表 4.6 に示す. 表内では学習済みモデル 1 をモデル 1, 学習済みモデル 2 をモデル 2 とした. 表 4.6 について, “No”はキーポイントのラベル名を表す.

結果として, 有意水準 5% で以下のように言える.

キーポイント No.7 モデル 2 のほうが優れている

Table 4.5: F 検定の結果

座標軸	有意差ありのキーポイント (No)	有意差なしのキーポイント (No)
x	2, 7, 8	1, 3, 4, 5, 6
y	1, 2, 3, 4, 5, 7	6, 8

Table 4.6: IS 無 (モデル 1) と IS 有 (モデル 2) の平均値, 標準偏差, 有意差

No	座標軸	モデル 1 の平均値	モデル 1 の 標準偏差	モデル 2 の平均値	モデル 2 の 標準偏差	平均値の 有意差
1	x	1.02	19.77	3.57	24.32	無
	y	51.81	121.05	36.27	80.08	無
2	x	3.38	18.09	2.65	10.94	無
	y	7.96	55.64	-0.43	40.93	無
3	x	2.77	14.32	0.26	13.89	無
	y	1.41	57.66	2.31	34.11	無
4	x	0.50	16.40	-3.71	17.82	無
	y	-43.26	147.56	-20.98	109.98	無
5	x	1.55	23.43	-4.69	28.94	無
	y	-21.63	113.35	-44.19	150.42	無
6	x	-3.20	20.32	-0.86	16.84	無
	y	8.55	97.31	10.63	88.28	無
7	x	3.72	15.87	0.43	7.16	無
	y	24.36	97.88	-1.57	42.45	有
8	x	1.87	20.53	0.95	13.17	無
	y	61.82	132.58	55.03	135.50	無

のことより、有意水準 5% でインスタンスセグメンテーションを行ったモデルの方がキーポイントの検出精度が優れている可能性が示された。

考察

検定では、有意水準 5% でインスタンスセグメンテーション学習を行ったモデルの方がキーポイントの検出精度が優れている可能性がある結果となった。マルチタスク学習はモ

モデル性能が高くなることが知られている。また、セグメンテーションタスクの内容に関してもキーポイント検出と関連性が高いため、このような結果になったと考えられる。また、この結果は He らの研究と同様の傾向を示す結果である。

4.2.3 実験 2：マスト，ブーム，リーチを 1 ネットワークで学習するか 3 ネットワークで学習するかの比較

実験 1 では、8 点のキーポイントをメインセールという 1 つのクラスが有するとしたが、ここでは 8 点のキーポイントをマスト、ブーム、リーチの 3 つのクラスが表 4.7 のように有するとして学習を行った場合にキーポイント検出に変化が見られるか比較を行う。この場合、セグメンテーション対象もメインセール全体ではなく、セール境界であるマスト、ブーム、リーチのみとなる。

クラス ラベル	保持するキーポイン ト (No)
マスト	1, 2, 3
ブーム	3, 4
リーチ	4, 5, 6, 7, 8

Table 4.7: クラスが持つキーポイント (再掲)

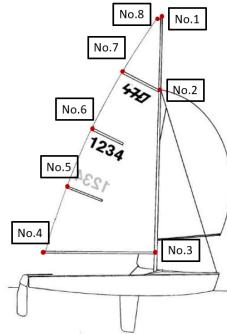


Fig.4.11: 検出するキーポイント (文献 [3] の図を改変)(再掲)

本実験では、その比較の予備実験として 3 クラスにてキーポイントを検出する場合、全てのクラスを 1 つのネットワークにて学習したモデルと、クラス 1 つごとに 1 つのネットワークを用いて学習したモデルに優劣を付けることを目的とした。

比較するモデル

本実験にて比較するモデルは、4.2.1 節にて述べた Train2 のマスト、ブーム、リーチクラスそれぞれを使って、ネットワーク ISC1×3 ネットワークを学習したものからなるモデル (モデル 3) と、Train2 によるネットワーク ISC3 を学習したモデル (モデル 4) である。モデル 3 では、それぞれがマスト、ブーム、リーチを 1 クラスずつ検出するのに対し、モデル 4 はマスト、ブーム、リーチの 3 クラスを 1 つのネットワークで検出する。

図 4.12 はモデル 3 での、図 4.13 はモデル 4 でのキーポイント検出器の損失関数の値の推移である。学習途中で検証を行っていないため、学習途中での学習不足や過学習の可能性は否めないが、損失関数の値の変化が少ない領域に達していると言える。

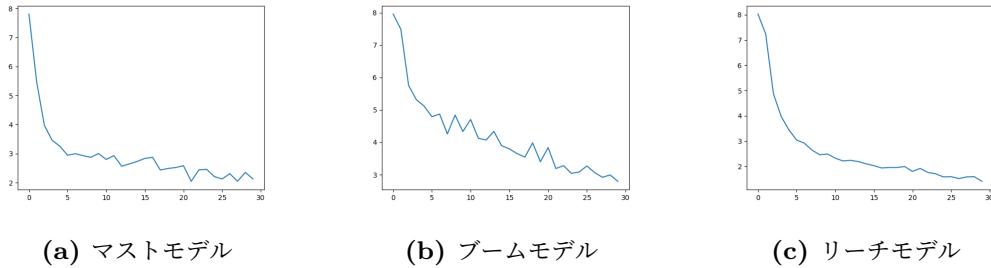


Fig.4.12: ネットワーク ISC1 にて学習したモデルの損失関数の推移

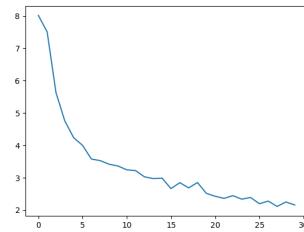


Fig.4.13: ネットワーク ISC3 にて学習したモデルの損失関数値の推移

結果

ネットワーク ISC1×3 によるモデル 3 の出力例は図 4.14、ネットワーク ISC3 によるモデル 4 の出力例は図 4.15 のようになる。

2 つのモデルが持つキーポイント検出結果を用いて検定を行った。ここでは、2 つのモデルの検出結果と真値との差の分布は正規性を持つと仮定している。

マストクラス、ブームクラス、リーチクラスの 3 クラスでの結果を示す。

マストクラスにおいて、78 枚のテストデータセットにおいて閾値より大きい confidence score を持つ出力が確認できたのは、モデル 3 にて 69 枚、モデル 4 にて 73 枚、であった。ここで、モデル 3 にてキーポイント検出を行った際の x 軸、y 軸における真値との差を正規化した数値の分布は図 4.16、モデル 4 では図 4.17 のようになった。図より、これらの分布は正規分布の特性を示唆していることが確認できた。

F 検定の結果を表 4.8 に示す。この結果を元に t 検定を行った。



(a) マストクラス

(b) ブームクラス

(c) リーチクラス

Fig.4.14: ネットワーク ISC1×3 ネットワークを学習したモデルの出力例

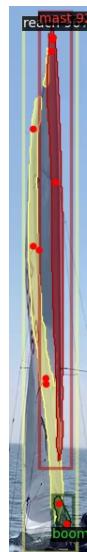


Fig.4.15: ネットワーク ISC3 を学習したモデルの出力

Table 4.8: F 検定の結果

座標軸	有意差ありのキーポイント (No)	有意差なしのキーポイント (No)
x	1, 3	2
y	1, 2	3

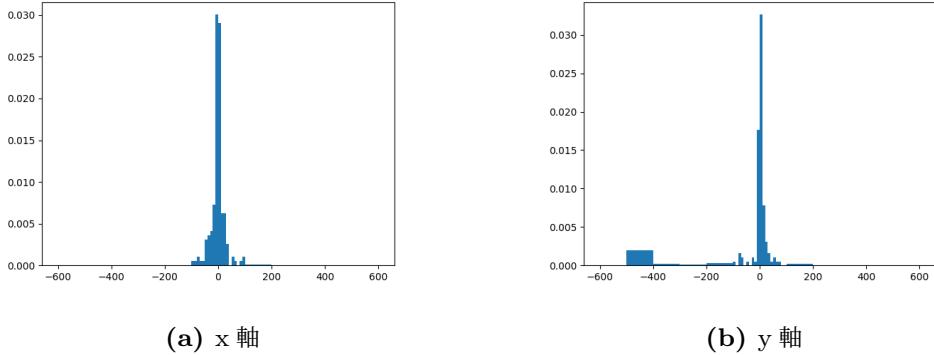


Fig.4.16: ネットワーク ISC1 にて学習したモデルのマストクラスの真値との差の分布

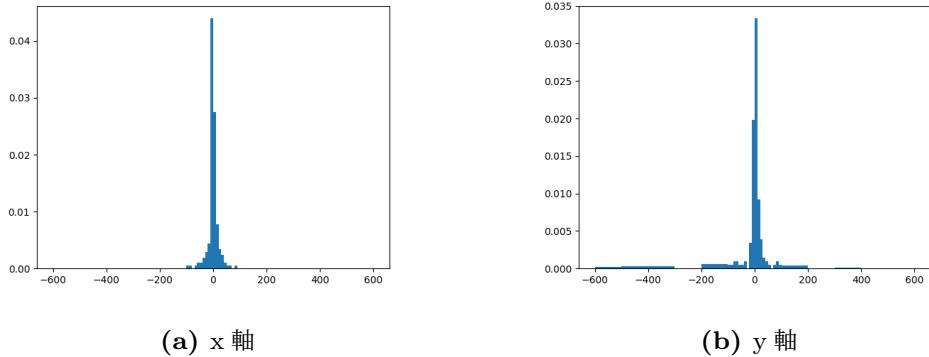


Fig.4.17: ネットワーク ISC3 にて学習したモデルのマストクラスの真値との差の分布

t 検定の結果を表 4.9 に示す。有意水準 5% において、2 つの条件で学習したモデルのキーポイント検出の精度には、キーポイント 3 の y 座標の検出に有意な差が見られた。このことより、有意水準 5% においてモデル 3 のほうが優れている可能性がある。

次に、ブームクラスにおいて、78 枚のテストデータセットにおいて閾値より大きい confidence score を持つ出力が確認できたのは、モデル 3 にて 40 枚、モデル 4 にて 59 枚であった。ここで、モデル 3 にてキーポイント検出を行った際の x 軸、 y 軸における真値との差を正規化した数値の分布は図 4.18、モデル 4 では図 4.19 のようになった。図より、これらの分布は正規分布の特性を示唆していることが確認できた。

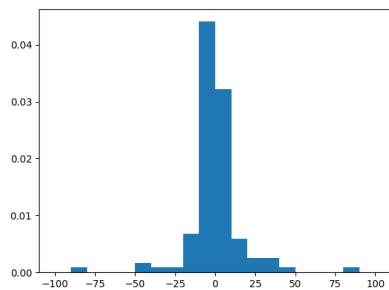
F 検定の結果を表 4.10 に示す。この結果を元に t 検定を行った。

t 検定の結果を表 4.11 に示す。有意水準 5% において、2 つの条件で学習したモデルのキーポイント検出の精度には有意な差があるとは言えない。

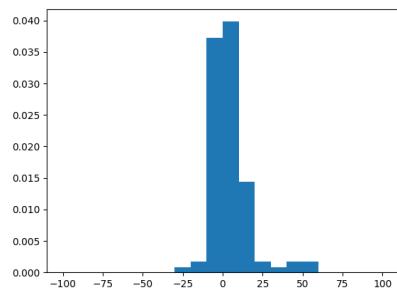
最後に、リーチクラスにおいて、78 枚のテストデータセットにおいて閾値より大きい

Table 4.9: モデル3 (ISC1×3) とモデル4 (ISC3) の平均値, 標準偏差, 有意差

No	座標軸	モデル3 の平均値	モデル3の 標準偏差	モデル4 の平均値	モデル4の 標準偏差	平均値の 有意差
1	x	-0.06	15.60	-1.02	8.84	無
	y	27.20	78.16	15.71	35.98	無
2	x	1.55	17.84	-1.50	17.68	無
	y	1.15	77.32	-10.05	59.98	無
3	x	-2.09	23.31	1.30	41.23	無
	y	-120.09	188.42	-287.01	228.37	有

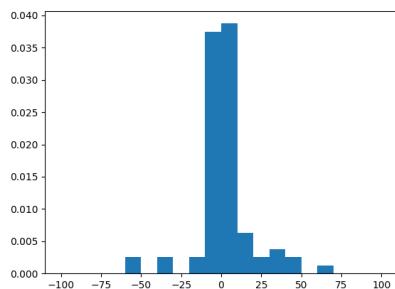


(a) x 軸

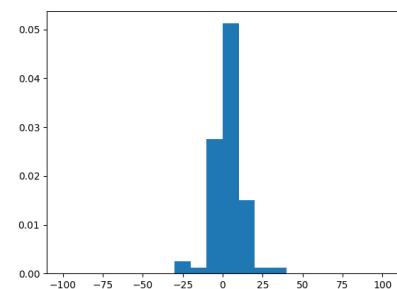


(b) y 軸

Fig.4.18: ネットワーク ISC1 にて学習したモデルのブームクラスの真値との差の分布



(a) x 軸



(b) y 軸

Fig.4.19: ネットワーク ISC3 にて学習したモデルのブームクラスの真値との差の分布

Table 4.10: F 検定の結果

座標軸	有意差ありのキーポイント (No)	有意差なしのキーポイント (No)
x	3, 4	
y	4	3

Table 4.11: モデル 3 (ISC1×3) とモデル 4 (ISC3) の平均値, 標準偏差, 有意差

No	座標軸	モデル 3 の平均値	モデル 3 の標準偏差	モデル 4 の平均値	モデル 4 の標準偏差	平均値の有意差
3	x	1.82	15.06	0.09	15.82	無
	y	6.66	9.06	9.28	13.45	無
4	x	2.19	18.98	0.12	16.39	無
	y	-0.04	6.37	-0.59	5.13	無

confidence score を持つ出力が確認できたのは、モデル 3 にて 77 枚、モデル 4 にて 77 枚、であった。ここで、モデル 3 にてキーポイント検出を行った際の x 軸、y 軸における真値との差を正規化した数値の分布は図 4.20、モデル 4 では図 4.21 のようになった。図より、これらの分布は正規分布の特性を示唆していることが確認できた。

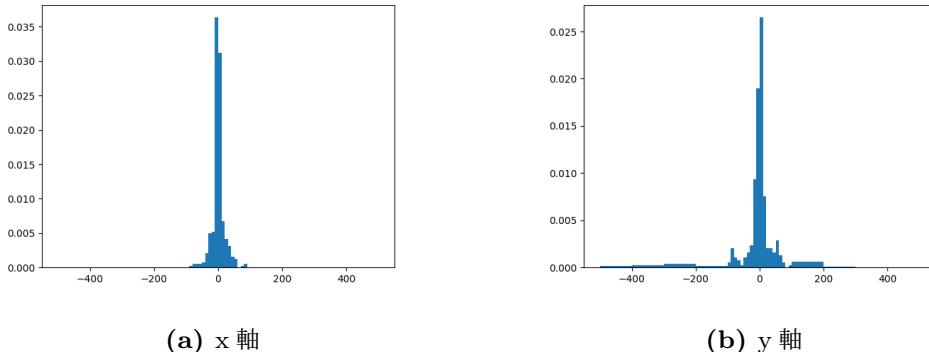


Fig.4.20: ネットワーク ISC1 にて学習したモデルのリーチクラスの真値との差の分布

F 検定の結果を表 4.12 に示す。この結果を元に t 検定を行った。

t 検出の結果を表 4.13 に示す。有意水準 5%において、2 つの条件で学習したモデルのキーポイント検出の精度には、キーポイント 5 の x 座標の検出に有意な差が見られた。こ

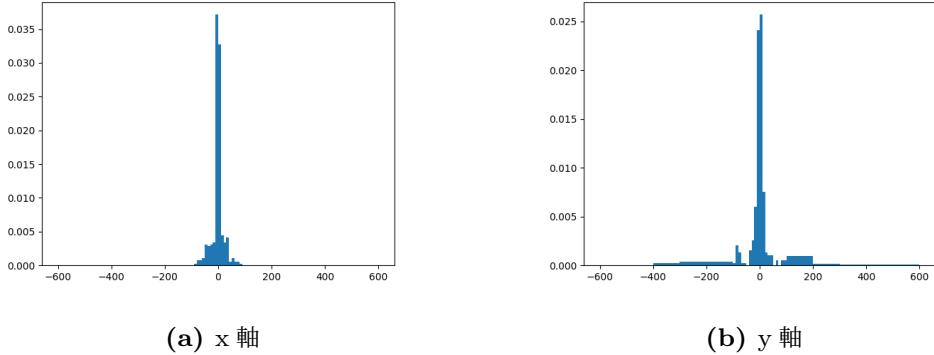


Fig.4.21: ネットワーク ISC3 にて学習したモデルのリーチクラスの真値との差の分布

Table 4.12: F 検定の結果

座標軸	有意差ありのキーポイント (No)	有意差なしのキーポイント (No)
x	4, 5, 6, 8	7
y	4, 6, 7, 8	5

のことより、有意水準 5% においてモデル 4 のほうが優れているといえる可能性がある。
結果として、それぞれのキーポイントについて、有意水準 5% で以下のように言える。

キーポイント No.3 モデル 3 のほうが優れている

キーポイント No.5 モデル 4 のほうが優れている

また、表 4.8, 4.12 の F 検定の結果について、有意差が見られたキーポイントの標準偏差は、表 4.9, 4.13 よりキーポイント 3, 4 以外はモデル 4 のほうが小さかった。confidence score が閾値より大きかった出力も全てのクラスにおいてモデル 4 のほうが多かったという結果となった。

考察

メインセール境界のマスト、ブーム、リーチの 3 クラスについて、ネットワーク ISC1×3 モデル（モデル 3）とネットワーク ISC3 モデル（モデル 4）の比較では、検出精度はどちらが優れているとも言えないが、検出数は全てのクラスでモデル 4 の方が優れていたため、モデル 4 の方が優れていると言えるのではないかと考える。これは、マスト、ブーム、リーチは相互に隣接しているため、他のクラスの検出結果が情報として利用できてい

Table 4.13: モデル 3 (ISC1×3) とモデル 4 (ISC3) の平均値, 標準偏差, 有意差

No	座標軸	モデル 3 の平均値	モデル 3 の 標準偏差	モデル 4 の平均値	モデル 4 の 標準偏差	平均値の 有意差
4	x	1.68	17.20	-3.15	24.71	無
	y	-11.24	81.08	-38.70	130.38	無
5	x	-8.67	34.34	3.21	28.10	有
	y	-40.93	169.71	-83.90	155.30	無
6	x	-0.20	17.74	-1.83	13.18	無
	y	-8.28	103.02	-18.04	69.35	無
7	x	0.94	12.61	3.25	15.02	無
	y	16.60	100.89	17.73	43.81	無
8	x	1.81	15.03	0.60	9.61	無
	y	43.85	93.37	31.44	61.96	無

る可能性が考えられる。

4.2.4 実験 3：物体領域についてメインセールとマスト, ブーム, リーチとの比較

本実験では, 8 点のキーポイントをメインセールという 1 つのクラスが有するとして学習したモデルと, マスト, ブーム, リーチの 3 クラスがそれぞれ有するとして学習したモデルの比較を行うことを目的とした.

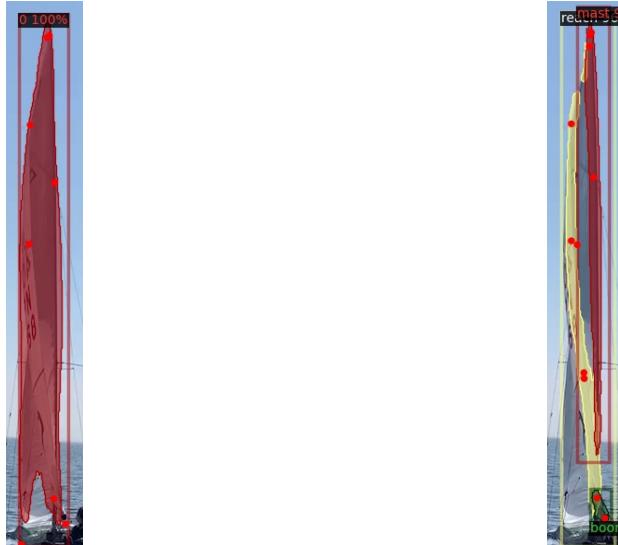
比較するモデル

4.2.2 節, 4.2.3 節の結果から, メインセール領域に基づくモデル 2 (ISC1/Train1) とマスト, ブーム, リーチ領域に基づくモデル 4 (ISC3/Train2) の比較を行った.

結果

メインセール領域に基づくモデル 2 による出力例は図 4.22a, マスト, ブーム, リーチ領域に基づくモデル 4 による出力例は図 4.22b のようになった.

2 つのモデルが持つキーポイント検出結果を用いて検定を行った. ここでは, 2 つのモデルの検出結果と真値との差の分布は正規性を持つと仮定おり, その分布はモデル 2 は



(a) メインセール領域に基づくモデル

(b) マスト, ブーム, リーチ領域に基づくモデル

Fig.4.22: 実際の出力例

4.2.2 節, モデル 4 は 4.2.3 節で示している.

マストクラスが持つキーポイント No.1, 2, 3, ブームクラスが持つキーポイント No.3, 4, リーチクラスが持つキーポイント No.4, 5, 6, 7, 8 にて結果を示す.

まず, キーポイント No.1, 2, 3 において, 78 セットのテストデータセットにおいて閾値より大きい confidence score を持つ出力を得たのは, モデル 2 にて 78 セット, モデル 4 にて 73 セットであった.

F 検定の結果を表 4.14 に示す. この結果を元に t 検定を行った.

Table 4.14: F 検定の結果

座標軸	有意差ありのキーポイント (No)	有意差なしのキーポイント (No)
x	1, 2, 3	
y	1, 2	3

t 検定の結果を表 4.15 に示す. 有意水準 5% において, 2 つの条件で学習したモデルのキーポイント検出の精度には, キーポイント 1 の y 座標, キーポイント 3 の y 座標の検出に有意な差が見られた. このことより, 有意水準 5% においてキーポイント 1 においてはマスト, ブーム, リーチ領域に基づくモデル 4, キーポイント 3 においてはメインセール領域に基づくモデル 2 のほうが優れている可能性がある.

Table 4.15: メインセールモデル 2 とマスト, ブーム, リーチモデル 4 の平均値, 標準偏差, 有意差

No	座標軸	モデル 2 の平均値	モデル 2 の標準偏差	モデル 4 の平均値	モデル 4 の標準偏差	平均値の有意差
1	x	3.57	24.32	-1.02	8.84	無
	y	36.27	80.08	15.71	35.98	有
2	x	2.65	10.94	-1.50	17.68	無
	y	-0.43	40.93	-10.05	59.98	無
3	x	0.26	13.89	1.30	41.23	無
	y	2.31	34.11	-287.01	228.37	有

次に, キーポイント No.3, 4 において, 78 セットのテストデータセットにおいて閾値より大きい confidence score を持つ出力が確認できたのは, モデル 2 にて 78 セット, モデル 4 にて 59 セットであった.

F 検定の結果を表 4.16 に示す. この結果を元に t 検定を行った.

Table 4.16: F 検定の結果

座標軸	有意差ありのキーポイント (No)	有意差なしのキーポイント (No)
x		3, 4
y	3, 4	

t 検定の結果を表 4.17 に示す. 有意水準 5% において, 2 つの条件で学習したモデルのキーポイント検出の精度には有意な差があるとは言えない.

最後に, キーポイント No.4, 5, 6, 7, 8 において, 78 セットのテストデータセットにおいて閾値より大きい confidence score を持つ出力が確認できたのは, モデル 2 にて 78 セット, モデル 6 にて 77 セットであった.

F 検定の結果を表 4.18 に示す. この結果を元に t 検定を行った.

t 検定の結果を表 4.19 に示す. 有意水準 5% において, 2 つの条件で学習したモデルのキーポイント検出の精度には, キーポイント 6 の y 座標, キーポイント 7 の y 座標の検出に有意な差が見られた. このことより, 有意水準 5% においてメインセール領域に基づくモデル 2 のほうが優れているといえる.

結果として, それぞれのキーポイントについて, 有意水準 5% で以下のように言える可

Table 4.17: メインセールモデル 2 とマスト, ブーム, リーチモデル 4 の平均値, 標準偏差, 有意差

No	座標軸	モデル 2 の平均値	モデル 2 の標準偏差	モデル 4 の平均値	モデル 4 の標準偏差	平均値の有意差
3	x	0.26	13.89	0.09	15.82	無
	y	2.31	34.11	9.28	13.45	無
4	x	-3.71	17.82	0.12	16.39	無
	y	-20.98	109.98	-0.59	5.13	無

Table 4.18: F 検定の結果

座標軸	有意差ありのキーポイント (No)	有意差なしのキーポイント (No)
x	4, 6, 7, 8	5
y	6, 8	4, 5, 7

Table 4.19: メインセールモデル 2 とマスト, ブーム, リーチモデル 4 の平均値, 標準偏差, 有意差

No	座標軸	モデル 2 の平均値	モデル 2 の標準偏差	モデル 4 の平均値	モデル 4 の標準偏差	平均値の有意差
4	x	-3.71	17.82	-3.15	24.71	無
	y	-20.98	109.98	-38.70	130.38	無
5	x	-4.69	28.94	3.21	28.10	無
	y	-44.19	150.42	-83.90	155.30	無
6	x	-0.86	16.84	-1.83	13.18	無
	y	10.63	88.28	-18.04	69.35	有
7	x	0.43	7.16	3.25	15.01	無
	y	-1.57	42.45	17.73	43.81	有
8	x	0.95	13.17	0.60	9.61	無
	y	55.02	135.50	31.44	61.96	無

能性がある。

- キーポイント No.1 モデル 4 のほうが優れている
- キーポイント No.3 モデル 2 のほうが優れている
- キーポイント No.6 モデル 2 のほうが優れている
- キーポイント No.7 モデル 2 のほうが優れている

考察

キーポイント検出の精度としてはモデル 2 の方が優れているキーポイント数が多い結果となった。モデル 4 の方が優れているとなったキーポイント 1 については、いずれのモデルにおいても平均値は全体的に低い数値を取っているため、メインセール領域に基づくモデル 2 のほうが優れていると言えるのではないかと考える。この理由として、メインセールは大きい物体であるため、検出が容易であるが、マスト、ブーム、リーチはメインセールの一部より、メインセールに比べると検出は困難であることが考えられる。

4.2.5 データ拡張学習に対する検定

本研究の問題点の一つは扱うデータセット数の少なさにある。本実験では、モデル学習の際にデータ拡張手法により学習用データセットを増加させることでキーポイントの検出に変化が生じるかの確認を目的とした。

比較するモデル

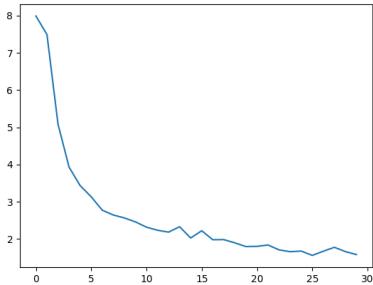
本実験にて比較する学習済みモデルは、メインセール領域セグメンテーションに基づいたモデルであり、4.2.2 節にて述べたモデル 2 と 4.2.1 説にて述べたデータ拡張データセット Train1a をネットワーク ISC1 にて学習したものである。これをモデル 5 とする。

以下の図 4.23 は 2 つの条件での学習におけるキーポイント検出器の損失関数の値の推移である。学習途中で検証を行っていないため、学習途中での学習不足や過学習の可能性は否めないが、図 4.23 より、損失関数の値の変化が少ない領域に達していると言える。

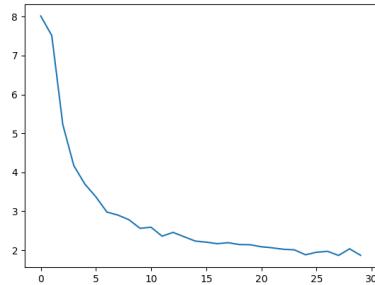
結果

データ拡張無しのモデル 2 による出力例は図 4.24a データ拡張有りのモデル 5 による出力例は図 4.24b のようになった。

2 つのモデルが持つキーポイント検出結果を用いて検定を行った。ここでは、2 つのモ

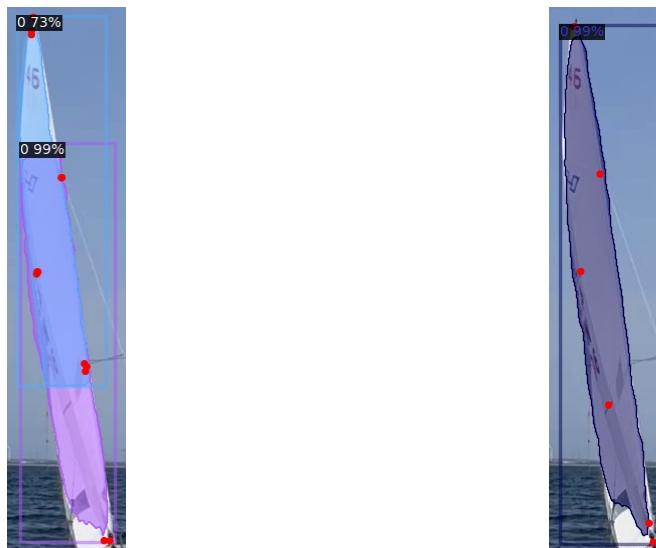


(a) データ拡張無しモデル



(b) データ拡張有りモデル

Fig.4.23: 損失関数の値の推移



(a) データ拡張無しモデル

(b) データ拡張有りモデル

Fig.4.24: 実際の出力例

デルの検出結果と真値との差の分布は正規性を持つと仮定している。モデル 2 にてキーポイント検出を行った際の x 軸, y 軸における真値との差を正規化した数値の分布は 4.2.2 に示した図 4.10 の様になり, モデル 5 では図 4.25 のようになった。図より, これらの分布は正規分布の特性を示唆していることが確認できた。

検出数について, 78 枚のテストデータの全てにおいて閾値より大きい confidence score を持つ出力が確認された。

F 検定の結果を表 4.20 に示す。この結果を元に t 検定を行った。

t 検定の結果を表 4.21 に示す。有意水準 5% において, 2 つの条件で学習したモデルの

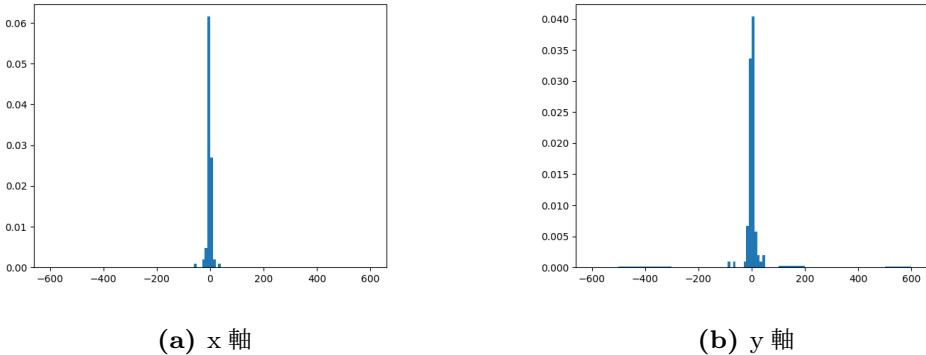


Fig.4.25: データ拡張有りモデルの真値との差の分布

Table 4.20: F 検定の結果

座標軸	有意差ありのキーポイント (No)	有意差なしのキーポイント (No)
x	1, 2, 3, 4, 5, 7, 8	6
y	2, 3, 4, 5, 6, 7, 8	1

キーポイント検出の精度には、キーポイント 2 の x 座標、キーポイント 3 の y 座標、キーポイント 5 の y 座標、キーポイント 7 の y 座標の検出に有意な差が見られた。

のことより、有意水準 5% で以下のように言える。

- キーポイント No.2 モデル 5 のほうが優れている
- キーポイント No.3 モデル 2 のほうが優れている
- キーポイント No.5 モデル 5 のほうが優れている
- キーポイント No.7 モデル 2 のほうが優れている

考察

データ拡張手法の適用は扱うデータセット数が少ないモデルの精度向上には有用な方法であったが、データ拡張手法を適用したのにもかかわらず、適用前に比べ全体的に精度の向上は見られなかった。これは、学習中に汎化性を保つための検証を行っていないことによる学習不足や過学習が発生している可能性がある。また、データ拡張手法を適用することによって、現実的ではない画像データを生成した可能性や、データ拡張手法に RandomLighting 関数を適用しているため、照明条件の悪い画像に対してはモデル 5 の

Table 4.21: データ拡張無しのモデル 2 と有りのモデル 5 の平均値, 標準偏差, 有意差

No	座標軸	モデル 2 の平均値	モデル 2 の標準偏差	モデル 5 の平均値	モデル 5 の標準偏差	平均値の有意差
1	x	3.57	24.32	0.77	16.68	無
	y	36.27	80.08	42.52	85.91	無
2	x	2.65	10.94	-1.23	5.24	有
	y	-0.43	40.93	0.52	13.31	無
3	x	0.26	13.89	-1.41	19.66	無
	y	2.31	34.11	-23.97	102.85	有
4	x	-3.71	17.82	-0.72	6.82	無
	y	-20.98	109.98	2.64	6.61	無
5	x	-4.69	28.94	-6.66	14.63	無
	y	-44.19	150.42	-0.54	82.50	有
6	x	-0.86	16.84	-5.77	19.20	無
	y	10.63	88.28	1.02	60.94	無
7	x	0.43	7.16	-4.02	29.44	無
	y	-1.57	42.45	21.36	79.99	有
8	x	0.95	13.17	-2.00	17.69	無
	y	55.03	135.50	27.07	63.21	無

方が優れた精度を持つ可能性も考えられる。

4.3 メインセールの形状可視化の IoU 比較実験

4.3.1 実験方法

ここでは detectron2 内のライブラリに含まれる Mask R-CNN により検出されたメインセールのキーポイント座標を元に, 提案システムにおけるメインセールの形状の可視化処理のアルゴリズムの実装を行い, その出力を確認した. 使用する Mask R-CNN のモデルは 4.2.1 節にて述べたモデル 2 とモデル 5 とした. 3.4 節にて説明を行ったアルゴリズムに従って, キーポイントの真値と検出座標を用いて, 内側面の領域を算出し, その 2 つの領域の IoU を求めた. 使用するテストデータセットは Mask R-CNN のテストデータ

セットと同様の 13 編の動画から 1 枚ずつ画像を抽出し、13 枚のテストデータセットを作成した。また、Mask R-CNN を利用したテストデータセットをデータ拡張なしで利用した。

4.3.2 結果

求めた IoU 結果を表 4.22 に示す。ここで、‘-’となっている部分は IoU の算出ができないものである（キーポイント検出が正しくない場合に生じる）。

Table 4.22: Mask R-CNN における内側面可視化の IoU の精度

テストデータセット	モデル 2 の IoU	モデル 5 の IoU
No.0	0.14	-
No.1	-	0.82
No.2	0.73	0.84
No.3	0.91	-
No.4	0.72	0.76
No.5	0.48	0.75
No.6	0.78	0.67
No.7	0.24	0.77
No.8	0.25	0.80
No.9	-	0.77
No.10	-	-
No.11	0.68	-
No.12	0.89	0.86
平均値	0.58	0.78
中央値	0.70	0.77

IoU の平均値、中央値はいずれもモデル 5 のほうが高い。また両方のモデルで IoU が算出できたテストデータに関して、ほとんどモデル 5 の方がより高く出ている。さらに中央値から極端に離れた値を示すデータセットはモデル 2 では 3 ないし 4 点存在するが、モデル 5 では見られない。このため、IoU を用いた比較では Mask R-CNN のモデルはデータ拡張学習無しのインスタンスセグメンテーション学習有りのモデル（図 2.8 のネットワークのもの）より、データ拡張学習有りのインスタンスセグメンテーション学習有り

のモデルの方が優れているといえると考える。

IoU を用いた比較は客観評価である。本来であれば被験者を用いた主観評価を行うべきであるが、今回は行わなかった。個人的な評価としては、全体的にデータ拡張学習有りのインスタンスセグメンテーション学習有りのモデルの結果は形状可視化を行った際の自然さという点で優れていると考える。個人的にモデル 5 がモデル 2 よりも優れていると感じたメインセール内側面の出力例は、図 4.26 である。

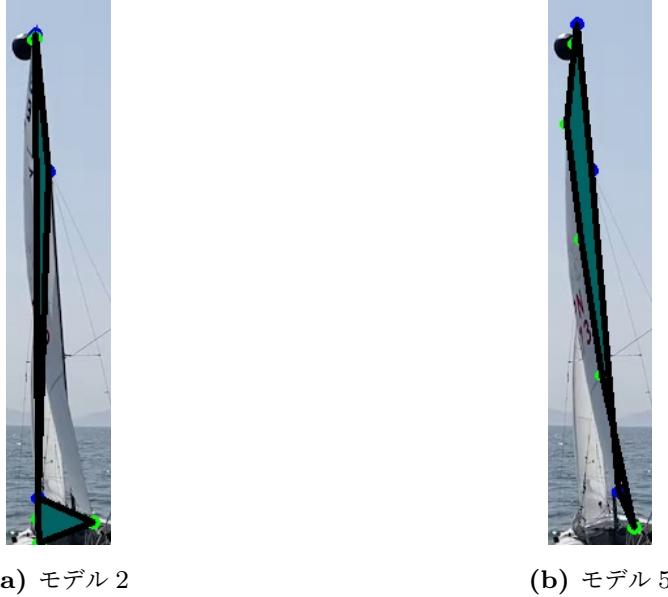


Fig.4.26: メインセール内側面の出力例 (No.8)

考察

本実験では、モデル 2 とモデル 5 について、メインセールの内側面の領域と真値の IoU を比較したところ、モデル 5 の方が優れた結果となった。4.2.5 節のモデル 2 とモデル 5 は t 検定、検出数にて優劣をつけることはできなかったが、データ拡張処理を行っているモデル 5 のキーポイントの検出座標の方が精度はよい可能性がある。また、両モデルのメインセールの内側面の出力を確認したところ、モデル 2 の出力はモデル 5 の出力に比べて、メインセールの外側にキーポイントが出力される事例が多くかった。これより、データ拡張有りで学習を行ったことにより、キーポイントの検出座標の真値との差の分布の分散について、IoU の値に大きく影響するキーポイントの分散が小さくなっている可能性がある。4.2.5 節の表 4.20, 4.21 より、 x , y 座標での F 検定に有意差が見られたキーポイントについて、キーポイント No.2, 4, 5 は x , y 座標でともにモデル 2 よりモデル 5 のほ

うが標準偏差が小さい。これより、メインセールの内側面の IoU について、キーポイント No.2, 4, 5 は大きな影響力を持つ可能性が考えられる。

4.4 実験結果のまとめ

4.2 節での比較より、有意水準 5% での検定や、検出数、IoU での比較の結果、4.2.1 節にて述べたモデル 5 が最も優れたモデルとなった。これより、本研究のキーポイント検出において、最も優れているモデルとされたのは、メインセール 1 クラスの外接矩形、ビットマスク、キーポイント座標をアノテーションし、さらにデータ拡張を行ったデータセットを用い、物体検出とインスタンスセグメンテーション、キーポイント検出の学習を行うネットワークモデルであった。これまで、真値のキーポイント 1 と 3 のユークリッド距離を 500 に正規化し、そのモデルでのそれぞれのキーポイントでの真値との差の平均値を求めてきたが、実際にこの距離を測定したところ、その距離は約 575cm であった。このことを踏まえ、表 4.23 においてモデル 5 でのそれぞれのキーポイントでの真値との差の平均値と実際の距離を示す。

また、4.3 説で使用した 13 枚のテストデータセットのモデル 5 による出力を図 4.27 に示す。

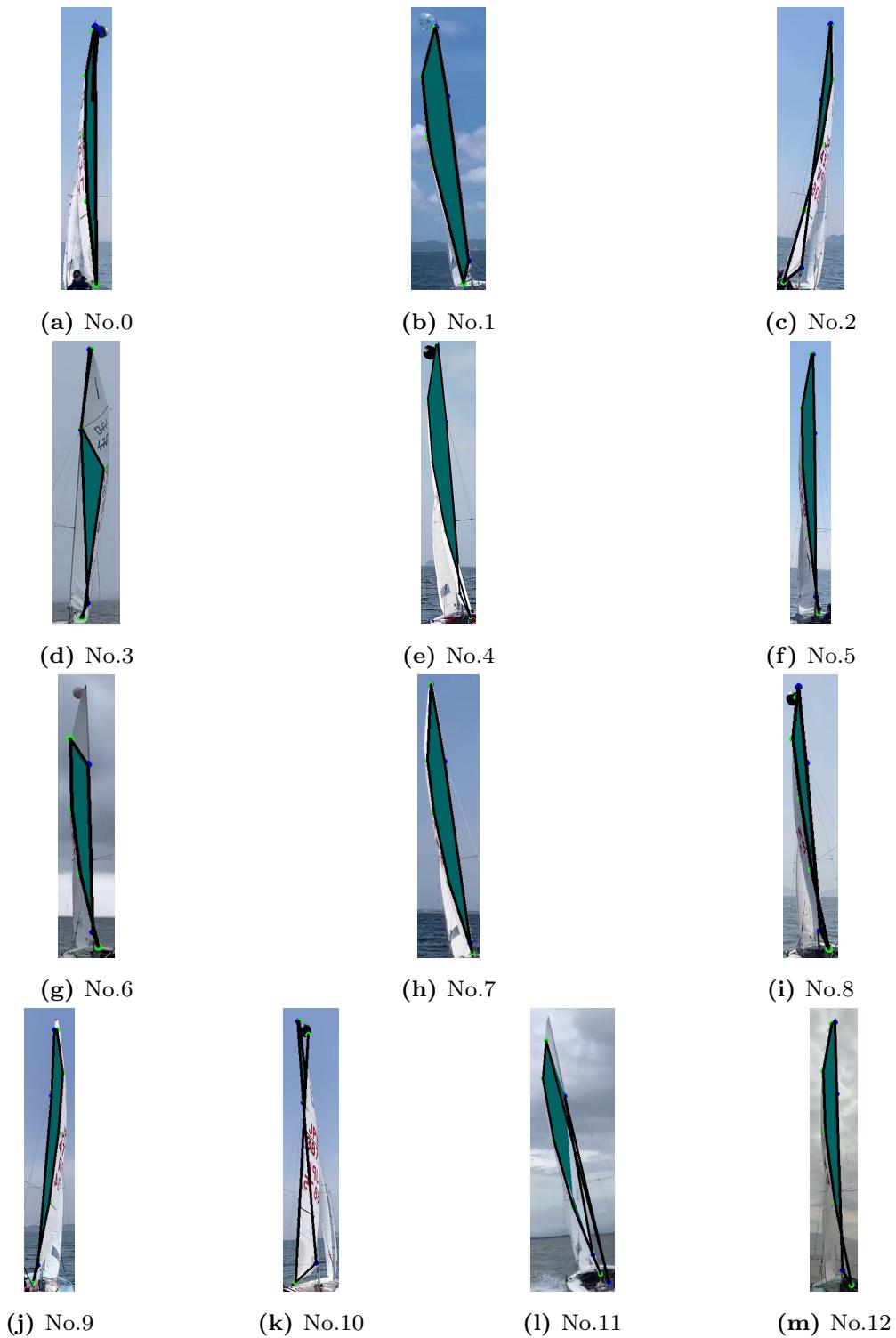


Fig.4.27: モデル5のメインセール内側面の出力例

Table 4.23: Mask R-CNN におけるキーポイント検出の精度

キーポイント番号	座標軸	平均値	実際の距離 (cm)
No.1	x	0.77	0.89
	y	42.52	48.90
No.2	x	-1.23	-1.41
	y	0.52	0.60
No.3	x	-1.41	-1.62
	y	-23.97	-27.57
No.4	x	-0.72	-0.83
	y	2.64	3.03
No.5	x	-6.66	-7.66
	y	-0.54	-0.62
No.6	x	-5.77	-6.64
	y	1.02	1.17
No.7	x	-4.02	-4.62
	y	21.36	24.56
No.8	x	-2.00	-2.30
	y	27.07	31.13

第 5 章

結論

5.1 まとめ

ヨット競技において、セールの形状は非常に重要な要素の一つであり、ヨット選手はセールの理想的な形状について練習時に議論を行うことで自身のセーリング技術の向上に努めている。そのためヨット競技における上達にはセールの形状を把握できることが大きく関係している。しかし、ヨット競技の初心者にとってこれは困難なことである。本研究では、ヨット初心者がヨットの練習を撮影した動画からのセールの形状の把握を手助けすることを目的として、携帯電話やスマートフォンに内蔵されたカメラで撮影されたヨットの海上練習の動画から、ヨットセールの輪郭や裏表の切り替わりを可視化するシステムを提案した。このシステムによって、ヨット初心者の早期上達を期待できる。ここで研究対象はヨットクラスの 1 つである 470 級クラスのメインセールとした。

今回提案したシステムは物体検出によりメインセールを検出し、その後キーポイント検出による特徴点の検出を行う。そこで検出されたキーポイント座標を用いてメインセールの輪郭と内側面の可視化を行う。システム内において物体検出手法としては YOLOv5、キーポイント検出手法としては Mask R-CNN を使用した。

YOLOv5 による物体検出の精度評価と Mask R-CNN によるキーポイント検出手法についてのモデルごとの比較を行った。YOLOv5 においては実際に検出した物体領域が真値を完全に含んでいるかを確認した。Mask R-CNN においては、複数の条件で学習を行い、それらのモデルの優劣を求めた。比較には有意水準 5% での F 検定・t 検定とモデルの物体検出数、メインセールの内側面とその真値との IoU を用いた。

結果として、YOLOv5 による物体検出の精度は、Recall 値が 0.91 と高い精度を得た。Mask R-CNN によるキーポイント検出については、最も優れているとされた学習条件は、

メインセールの外接矩形とインスタンスセグメンテーション、キーポイントの検出を学習し、その際に学習用データセットのデータ拡張処理を行うものであった。そのモデルにて、検出されたキーポイント座標の真値との差の平均値は実世界において x 軸方向（右を正）には -7.66cm から 0.89cm 、 y 軸方向（下を正）には -27.57cm から 48.90cm の差を持った。また、検出されたキーポイントを用いた内側面の着色領域と真値との IoU は平均で 0.78 であった。

本研究では、ヨットの海上練習の動画より抽出した画像データを用いて YOLOv5 の精度評価と Mask R-CNN のモデルごとの比較を行った。YOLOv5 の検出精度は高い精度を持ったが、Mask R-CNN のキーポイント検出はキーポイントごとの検出座標とその真値との差の平均値にはばらつきが存在した。このことから、今後はキーポイントの検出精度の向上を行うための検討が必要である。また、今後は画像や動画からメインセールの深さ推定の研究の基礎となることを期待する。

5.2 今後の課題

実際にヨット競技の初心者に向けたメインセールの可視化システムとして実現するためには、キーポイント検出の精度向上が課題となる。これに対する改善案としては、使用するデータセットの拡張が挙げられる。今回、キーポイント検出のデータセットとして 36 編の動画を用意し、23 編の動画から 148 枚の画像を抽出したが、同じ動画から抽出した画像は疑似した構造をしているため、学習モデルは 23 パターンに分類できる 148 枚の画像を学習したことになる。本研究ではその対策としてデータ拡張処理を行ったが、今後はヨットの動画数を拡張し、この課題を改善したい。このこと以外にも、キーポイント検出の精度の問題の一因として考えられるのは学習中に汎化性の検証を行っていないことである。そのことにより、過学習を起こしているかを判断できなかった。Mask R-CNN の学習に汎化性の検証と損失関数の推移によって学習を打ち切りにするアルゴリズムの追加も今後の課題である。キーポイント検出の精度が向上することで、マスト、ブーム、リーチの座標をより正確に求めることができ、メインセールの深さ推定を行う基礎となることが期待できる。また、より正確な輪郭の抽出のために、正確なキーポイント座標を用いることも考えられ、マーカーを取り付けることなく、セールの立体形状を抽出できるシステムの実現が期待できる。キーポイント間の幾何学的、物理的特徴を用いたセールの曲面による 3D 形状の復元も期待される。

謝辞

本研究を進めるにあたり適切なご指導とご助言を下さった名古屋大学大学院情報学研究科 工藤 博章 准教授、松本 哲也 助教、大同大学情報学部 竹内 義則 教授に深く感謝いたします。また、工藤研究室の皆様に心より感謝し、厚く御礼申し上げます。

参考文献

- [1] Maciel, L., Marroquim, R., Vieira, M. et al. “Monocular 3D reconstruction of sail flying shape using passive markers.” Machine Vision and Applications 32, 26 (2021).<https://doi.org/10.1007/s00138-020-01149-3>
- [2] “470 class four-seventy”. 日本 470 協会公式ウェブサイト.<https://470jpn.org/about470/index.html> (参照 2023-12-31)
- [3] “INTERNATIONAL 470 CLASS RULES 2017”. 日本 470 協会公式ウェブサイト.https://www.470jpn.org/_src/sc1777/ClassRules_2017_amend_j.pdf (参照 2023-12-31)
- [4] 梅田洋一. “ヨットを題材にした中高生向けの力学”. 物理教育 第 52 卷 第一号 p13-16.2004. https://www.jstage.jst.go.jp/article/pesj/52/1/52_KJ00005896918/_article/-char/ja/
- [5] Yang, G., Feng, W., Jin, J. et al. “Face Mask Recognition System with YOLOV5 Based on Image Recognition.” 2020 IEEE 6th International Conference on Computer and Communications(2020).<https://ieeexplore.ieee.org/document/9345042>
- [6] He, K., Zhang, X., Ren, S., Sun, J. “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition.” arXiv.2015-04-23.<https://arxiv.org/abs/1406.4729>
- [7] Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R. Detectron2. 2019. Available online: <https://github.com/facebookresearch/detectron2> (参照 2024-01-04).
- [8] He, K., Gkioxari, G., Dollár, P., Girshick, R., “Mask R-CNN”. In : ICCV(2017). 2017. pp 2961-2969. https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf

- [9] Ren, S., He, K., Girshick, R., Sun, J. “Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks” . NIPS 2015. 2015.
https://proceedings.neurips.cc/paper_files/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html
- [10] He, K., Zhang, X., Ren, S., Sun, J. “Deep Residual Learning for Image Recognition” . arXiv. 2015-12-10. https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf
- [11] COCO, <https://cocodataset.org/#home> (参照 2024-02-08)
- [12] labelImg, <https://github.com/HumanSignal/labelImg> (参照 2024-02-08)
- [13] coco-annotator, <https://github.com/jsbroks/coco-annotator> (参照 2024-02-08)