

TUMSAT-OACIS Repository - Tokyo

University of Marine Science and Technology

(東京海洋大学)

時系列データを用いた船舶認識に関する研究

メタデータ	言語: jpn 出版者: 公開日: 2018-12-17 キーワード (Ja): キーワード (En): 作成者: 張, 婷 メールアドレス: 所属:
URL	https://oacis.repo.nii.ac.jp/records/1649

修士学位論文

時系列データを用いた船舶認識に関する研究

平成 30 年度

(2018 年 9 月)

東京海洋大学大学院

海洋科学技術研究科

海運ロジスティクス専攻

張 婷

修士学位論文

時系列データを用いた船舶認識に関する研究

張
婷

2018.9

目次

1、	研究の背景	1
2、	関連研究	3
3、	研究の目的	5
4、	機械学習	6
5、	機械学習アルゴリズム	9
	5-1 教師あり学習	9
	5-2 教師なし学習	9
	5-3 強化学習	10
	5-4 汎用な機械学習アルゴリズム	10
6、	深層学習による物体認識	15
	6-1 深層学習	15
	6-2 畳み込みニューラルネットワーク	15
	6-4 深層学習フレームワーク	16
	6-5 深層学習手法	17
7、	航行中の船舶からの物体の認識	25
	7-1 データセット 1	25
	7-2 データセット 1 を用いた静止画像の認識結果	34
	7-3 データセット 1 を用いた動画の認識結果	40
	7-4 データセット 2	42
	7-5 データセット 2 を用いた静止画像の認識結果	46
	7-6 データセット 2 を用いた動画の認識結果	49
8、	実験	52
	8-1 実験環境	52
	8-2 航行中の船舶から取得した映像による実験	52

9、時間情報を利用した認識精度の向上	59
9-1RNN	59
9-2LSTM	62
9-3GRU	64
9-4QRNN	65
9-5LSTM、GRU に関する実験	66
9-6 SSD-LSTM に関する結果	67
10、評価	70
11、考察	71
12、まとめ	72
13、参考文献	73
謝辞	76

1、 研究の背景

20 世紀 90 年代以降、世界貿易の自由化、経済総合と情報グローバリゼーションの加速に伴い、世界の海運業界は大きく発展した。世界の商船数量は増加しており、船舶の規模はますます拡大している。海上輸送の発展は、国の経済を支えるだけでなく、ある程度経済発展を促進している。一方、海上輸送の増加に伴い、海上安全問題は海運問題の重要な一部になっている。海難事故による人命、財産の損失は数えきれない。同時に海難事故においては、海洋環境と海洋生物に直接的な危険をもたらす。そのため、海難事故の削減は全世界にとって大きな課題である。

船舶衝突、火災、爆発、船舶の紛失、船舶故障を引き起こす多くの種類の事故がある。衝突の原因としては気象条件、船舶の技術状態、操船者の技術レベル、港湾施設と管理レベルなど様々な原因がある。多くの自然条件や客観的な原因が存在するが、人為原因は主な原因である。ほとんどの事故は操船者の見張り不十分によって引き起こされた。海上保安庁[1]からの平成 29 年度海難情報によって、平成 29 年度に認知した船舶事故数は 1,977 隻で、そのうち 44%がプレジャーボート、27%が漁船、9%が貨物船である。平成 29 年の事故原因別の隻数は見張り不十分 358 隻で、過去 5 年間の事故原因の累計で人為原因が 7826 隻(74%)に占めている。その結果を見ると、「見張り不十分」が主な原因であることが分かっている。また、船舶事故により、死者や行方不明の人数は 82 人がいる。その以外に、船舶事故以外の乗船中の死者数および海浜事故の死者数は 2633 人がいる。平成 29 年において重大な海難事故が発生している。2 月 10 日青森県大間崎沖に転覆していた漁船が発見され、乗船者 2 名は死亡が確認された。7 月 26 日に旅客船が神戸空港東進入灯台の支柱に衝突したため、乗客 15 名は負傷した。海難事故による死傷者は毎年発生している。海難事故の削減のために、見張り支援の研究が不可欠と思う。

さらに、海上衝突予防法[2]第 5 条によって、船舶は周囲の状況および他船との衝突の恐れについて判断できるように、適当な手段で、常時な見張りをしなければならないと述べている。現在、様々な見張り支援手段が導入されている。2008 年以後、VHF 帯デジタル無線機器で受信したデータを電子海図やレーダー画面上に表示できるこの要件を満足する船舶には AIS を搭載する義務がある。AIS 装置を通じ、他船の船名、位置情報、速度、速度などの情報を全部取得し、海難の削減に大きな貢献をした。海上保安庁からの統計情報によって、近年日本の沿海海域に通航している船舶の中に、AIS が搭載された船舶は一日ほぼ 7000 隻に占めている。現在、ほとんどの船舶は AIS が搭載しているが、船舶事故がよく発生するプレジャーボート、漁船などの小型船には大部分 AIS が搭載していない。しかし、プレジャーボートや漁船などの小型船に対する見張り支援の研究はまだされていない。これより、コンテナ船や LNG 船などの大型船だけではな

く、様々な船舶やブイなどの障害物を認識、表示し、早めに操船者を知らせ、有効に避航することが必要である。

人工知能の発展に伴い、機械学習はさまざまな分野に応用されている。コンピュータ視覚、物体検出はそのうちのひとつである。現在、機械学習を用いた車の自動運転の技術も存在している。自動運転システムは車載感知センサー、レーダー、高精度地図と AI アルゴリズムなどを利用して障害物を検出し、自動運転を実現する。今自動運転は完全に道路交通に応用できないが、自律走行技術は成熟している。1960年代には、遠隔操作の無人艦隊が軍事分野で広く使用されていた。近年、自動制御、インターネット、遠隔操船などの急速な発展に伴い、船舶に関する環境認識技術や通信ナビゲーション技術も広く利用され、それは知能船、無人船の開発のための技術的実現可能性を提供した。そのため、車の自動運転と同じ、機械学習を用いて他船やブイなどの障害物を認識、表示することで、操船者を支援する同時に、自動操船の実現も期待できる。

本研究では、海難事故の削減ために、機械学習を利用して、船舶認識の見張り支援手法を提案した。

2、関連研究

見張り支援の研究として、高橋ら[3]が ORB 特徴量を用いて、テンプレートマッチングを行うことで、SW/HW 協調処理による船舶自動見張りシステムを提案している。このシステムでは見張りの自動化を実現するために、ORB 特徴量から抽出した特徴を FPGA に並列処理を行う。また、AIS 情報を用いた小形船用見張り支援システムも提案している。ORB 特徴量を用いた見張り支援システムは船舶の画像認識に実現できるが、FPGA に画像の処理時間がかかる欠点もある。また、丹羽ら[4]は GPS を用いて自船の動静と他船の動静ともに AIS で把握し、衝突の事前検出、警報の提示が行えるような支援システムのコンセプトを作成した。このシステムでは自船と他船両方が AIS を搭載し、両方の AIS 情報を通じ、CPA および OZT のアルゴリズムで衝突の危険性あるかどうかを判断する。AIS 情報を通じ、自船と他船両方の情報を全部取得できるが、実時的に航海映像の上で他船の位置情報と船舶種類を表示できない。それ以外、羽子田ら[5]は河川交通におけ、魚眼カメラを利用して河川の 3D 地図を作成する河川見張り支援手法を提案した。3次元地図で河川の水路と河川的环境状況が分かるため、有効に他船やブイなどの障害物を回避することが期待できるが、河川の上の障害物の具体的な位置情報は取得できない。

船舶認識の研究例はいろいろがある。稲石ら[6]は海上航行のための視覚支援ニューラルネットワークシステムを提案している。この論文では航路推定の目的とする 3層画像認識ニューラルネットワークモデルを作成した。実験用の三次元のコンピュータ船舶設計図は日本の代表的な大型船を用いて 10 種類を分類し、学習用データセットとする。3層ニューラルネットワークは入力層、隠れ層と出力層を分け、入力層は 36 ユニット、隠れ層は 108 ユニット、出力層は 26 ユニットがある。航路推定の正確な認識率と船舶認識率は、それぞれ 55.5%と 44.4%の結果を取得した。稲石らが提案した視覚支援ニューラルネットワークシステムは船舶認識と航路の方面に良い結果を取得し、操船者に支援することに大きな貢献した。また、近藤ら[7]は操船者を支援するために、機械学習に基づくニューラルネットワークや HOG (Histogram of Oriented Gradient) を用いて航海環境画像から障害物を分類する手法を提案した。この論文では河川交通の中から撮影された画像を 9つのカテゴリに分類し、HOG とニューラルネットワークを用いて画像を分類する。また、機械学習による障害物分類の結果を比較、検討した。それ以外、この論文は時系列情報を用いた障害物分類手法を提案した。実験の結果として、河川航行カメラを用いた障害物の分類実験の正解率は約 50%に得た。この二つの論文は機械学習に基づいて見張り支援に関する研究である。

時代の変遷に伴い、人工知能の発展はますます進んでいる。機械学習に基づく多くの画像認識アルゴリズムも存在している。機械学習を用いた画像認識の研究としては、R-CNN[8]に基づく物体認識の研究が提案されている。Ross Girshick らは Selective

Search[9]この手法を用いて、画像の候補領域を生成する。それぞれの候補領域をリサイズし、ニューラルネットワークにかけることで画像の中の物体を検出する。また、閾値 IOU この概念を導入された。最初に画像は入力値とし、続いて Selective Search 手法で似た領域を階層的にグルーピングして、約 2000 個程度の物体領域の候補を出す。Selective Search からの候補領域をリサイズし、ニューラルネットワークに輸送する。ニューラルネットワークからの feature map を SVM (IOU を用いたサンプル分類器で各カテゴリ分類する。最後、NMS 活性化関数(non-maximum suppression、非極大値抑制関数)で一番高いスコアの領域を表示する。この論文では IOU が 0.5 より大きい場合はその候補領域を positive とする。逆に、その候補領域を捨てる。この論文は CNN の softmax 分類器が使っていない。SVM 分類器を使ったため、精度が上がった。また、このモデルは各カテゴリに一つの分類器を生成するから、ネットワークは複雑である。

また、2016 年に Mask-CNN [10]この論文は M-CNN モデルを提案した。学習するときに part annotation と image-level ラベルこの二つの情報だけが入力される。そのうち part annotations は二つ部分を分け、head part と torso part である。M-CNN は 4 つのストリームモデルで、4 つの入力それぞれは元画像、頭の部分、体の部分、物体の部分である。それぞれの部分をニューラルネットワークかけ、deep descriptors あるいは feature map を取得し (Object feature, Torso feature, Head feature, Image feature)、次にこれらの 1024-d 特徴ベクトルを取得し、4 つのベクトルをつなぎ合わせ、最終的にカテゴリを 12 正則化関数、fc 層、softmax 損失関数で取得する。この論文では物体が頭の部分と体の部分はそれぞれ CNN をかけるため、M-CNN の認識精度 85.5% に達した。しかし、船舶認識に応用すると、データセットの作成は難しい。また、船舶の位置情報とラベル情報を画像で表示できない。

3、 研究の目的

海難事故により財産、人命の損失を見積もることができず、ある程度までは海洋環境や海洋生物に大きな影響を与える。海難事故の削減は船員の責任であるだけでなく、私たち一人ひとりには責任と義務がある。見張り不十分が海難事故の主な原因なため、見張り支援が不可欠である。現在様々な見張り支援手法があるが、プレジャーボート、漁船などの小型船とブイの障害物認識に関する研究を行っていない。海上航行中に、AIS が搭載された大型船の認識が容易であるが、AIS が搭載されていない小型船やブイに対する認識が非常に難しい。しかし、機械学習に基づく障害物認識は AIS と関係なく、船舶から撮影されて映像の中の障害物を認識する視覚支援手法である。この手法により、早めに他船やブイなどの障害物を認識、表示できる。近 50 年以來、情報技術の発展に伴い、物体検出の領域に巨大な成功を獲得した。特にニューラルネットワークを用いた画像認識手法に関して多くの研究者が報告している。現在、機械学習は様々な分野で活用され、画像認識の領域に大きな成就を取得した。本研究では海上に存在する船舶やブイなどの障害物を学習用画像とし、航行中の船舶から撮影された映像から障害物の識別、表示することが本研究の一つの目的である。また、障害物認識の精度の検討や複数手法の比較検討を行う。同時に障害物認識の精度を向上するために、時系列データを用いて、船舶認識モデルを検討した。時系列ニューラルネットワークは RNN に基づいて、時間序列があるデータを用いて、物体の発展と動向を予測できるネットワークである。

4、機械学習

機械学習 (Machine learning) は多領域の学科で、確率論、統計学、近似理論などの多くの学問分野を含む多分野の科目、人工知能の一部である[11]。学習は人間の本来の一つ能力である。名前が示すように、機械学習は、コンピュータに人と同じ学習能力を持たせることである。コンピュータが新しい知識やスキルを取得し、既存の知識構造を再編成してパフォーマンスを継続的に向上させるために、人間の学習行動をシミュレートまたは実装する方法を専門にする。機械学習の定義は最初 1996 年に Langley から提出した。機械学習は人工知能科学で、主な研究対象が人工知能、特に経験的学習における特定のアルゴリズムの性能を向上させる方法である。そのあと、1997 年に Tom Mitchell はもっと詳細な定義があり、機械学習は自分の経験を通じ、自動的に改進のコンピュータアルゴリズムであると言った。2004 年に機械学習の定義は変更された。Alpaydin は機械学習がデータと以前の経験を用いてコンピュータアルゴリズムを最適化させる性能標準と定義した。1959 年アメリカの研究者 Samuel は学習の能力を持たすチェスプログラムを設計した。このチェスプログラムはゲームプレイでスキルを学び、最後、開発者本人を敗北した。また、2016 年に google による開発した人工知能ロボット AlphaGo はチェス世界チャンピオンを敗北した。そのことで、人工知能は非常に前向きな価値があることがわかる。

伝統的には、コンピュータを動作させたい場合は、一系列命令をコンピュータに与えてから、手順をステップごとに実行する。しかし、機械学習は、入力したデータを受け入れるしかない。つまり、機械学習は、コンピュータが命令ではなくデータを使用してあらゆる種類の作業を行う手法である。「統計的」思考には常に「機械学習」という概念が付随し、機械学習が統計的アルゴリズムを学習するとも言える。実際に、機械学習の主な目的の1つは、人間の思考のプロセスを、コンピュータがデータを処理してモデルを計算するプロセスに変換することである。機械学習は、コンピュータが既存のデータあるいは経験を用いて、モデルを生成する手法であり、このモデルを用いて予測する方法である。このプロセスは以下の図1に示すようである。

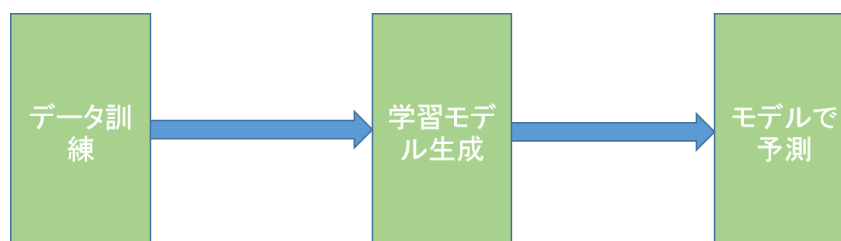


図1.機械学習のプロセス

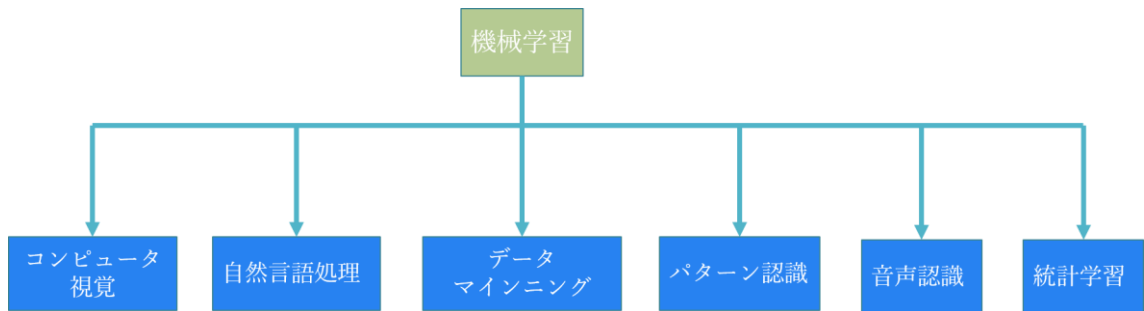


図2.機械学習の関連領域

機械学習は様々な分野に活用している。たとえば、模式識別、統計学習、データマイニング、コンピュータ視覚、音声認識、自然言語処理などの領域に深い関連性を持っている。機械学習がデータを学習する過程なため、一般的に、データマイニングは機械学習と同じにすることができる。上の図2は、機械学習に関連するいくつかの関連分野と研究分野を示している。

- パターン識別：パターン識別は機械学習に相当する。「Pattern Recognition And Machine Learning」この本で、Christopher M. Bishop は「パターン認識は工業界から来ており、機械学習はコンピュータ学科から来ている。しかし、彼らは同じ分野の2つの側面として見ることができ、過去10年間でかなり発展している」と言った。
- コンピュータ視覚：コンピュータ視覚は画像処理と機械学習の総合と見ることができる。画像を機械学習の入力とし、一連の処理を通じ、学習モデルを生成する。この学習モデルを用いて画像を認識する。コンピュータ視覚はいろいろな関連応用があり、たとえば、画像認識や手書き数字認識などの応用がある。
- データマイニング[12]：データマイニングは機械学習を活用して膨大なデータベースからデータあるいは知識を抽出する過程である。ただ、抽出のデータは三つの条件を満足しなければならない。その三つの条件は役に立つデータ、自明でないデータ、人間にも分かりやすいデータという条件である。データ抽出の種類により適用な手法も異なる。データマイニングの手法は四つがあり、それぞれは分類、特徴付け、相関、クラスタリングである。今まで、具体的な応用がまたない。
- 統計学習：上記の話したとおり、機械学習は統計学と関連性がある学科である。機械学習は統計学習に近似するとも言える。たとえば、機械学習の中の多数の分類アルゴリズム SVM(サポートベクターマシン)は統計学科の知識で非線形の決定境界を持つデータセットを分類する[13]。両者の区別は統計学習が統計模

型の発展と最適化に注目し、機械学習が問題の解決に注目する。

- 自然言語識別：自然言語はテキスト処理と機械学習の総合である。コンピュータに人間の言語を分析、理解する過程である。コンピュータの言語は二進法で、人間の言語は様々な言語があるため、自然言語識別はコンピュータに人間の言語の語彙や文法などを学習させて、我々の言語を分析、理解できるようなアルゴリズムである。
- 音声認識：音声認識は機械学習を用いた音声処理手法である。つまり、機械学習に基づくオーディオ処理技術である。一般に、音声認識は単独に使用されず、自然言語処理の関連技術と取り込んでいる。たとえば、Apple の音声助手 Siri は一つの例である。

5、機械学習アルゴリズム

第四章に機械学習の発展過程と応用は大体上分かったため、この章に機械学習の主なアルゴリズムを紹介する。全体として、機械学習アルゴリズムは3つのカテゴリ、教師あり学習、教師なし学習、強化学習に分けることができる。教師あり学習は入力と対応の出力を学習モデルであらかじめ学習し、学習済モデルであたらしいデータの出力を予測する。逆に、教師あり学習は入力と対応の出力を学習モデルであらかじめ学習し、学習済モデルであたらしいデータの出力を予測する。強化学習は入力と出力がなく、学習モデルが自分で周囲の環境を見て、どう行動すべきかを判断する。

5-1 教師あり学習

動作原理：アルゴリズムは独立変数（共変量、予測変数）と従属変数（結果変数）で構成され、独立変数から結果変数を予測する。これらの変数集合を通して、入力から出力に反射する関数を生成する。学習過程には、設定した損失閾値に達したら、学習を中止する。つまり、モデルが必要な精度のレベルに達したら、モデルの学習を終わりにする。学習のときに、訓練データがコンピュータに入力するだけでなく、分類の結果（データのラベル）も分析のためにコンピュータに入力する。

コンピュータが学習すると、新しい未知のデータが予測される。また、データがさまざまな結果の確率を計算し、最も正確な結果を得られる。つまり、一番高い確率の結果を出力する。

コンピュータは学習の過程で学習データを持っているだけでなく、訓練結果（ラベル）も持っているので、訓練効果は良好である。本研究で使った学習モデルも教師あり学習である。学習用船舶の静止画像と船舶のレベルは学習モデルに入力する。その後、学習済みモデルで新しい画像を認識する。

教師あり学習の例：回帰[14]、決定木[15]、ランダムフォレスト[16]、KNN（K近傍法）[17]、ロジスティック回帰[18]など。

5-2 教師なし学習

動作原理：教師なし学習アルゴリズムでは、予測のための目標変数または結果変数は用意していない。すなわち、コンピュータに学習データのみを入力する。結果（ラベル）はないので、コンピュータはどのデータがどのラベルを持っているかを正確に知ることができず、コンピュータ自身の計算能力でデータの特性を分析しかない。

教師なし学習の例：Apriori アルゴリズム[19]、K-means(K平均法)[20]。

5-3 強化学習

動作原理：強化学習、評価学習、学習とも呼ばれる強化学習は、単一の方法ではないが、機械学習の方法では、知能制御ロボットの分析と予測の分野で多くの応用がある。

強化学習の例：マルコフ決定[21]。

5-4 汎用な機械学習アルゴリズム

5-4-1 線形回帰

線形回帰(Linear Regression)は連続変数の推定に使用される。ここでは、独立変数と従属変数の間に最も適合する直線を当てはめ、それらの間の関係を確立する。この線は回帰線と呼ばれ、簡単に線形方程式： $y = a * x + b$ で表される。

上記の式において、それぞれの意味は以下の通りである。

y：従属変数

a：勾配

x：独立変数

b：インターセプトターム

5-4-2 ロジスティック回帰

予測対象が確率的である場合、値域は0以上1以下を満たす必要がある。確率は0から1の間であるので、0.5で閾値を定めると分類モデルを作成できる。ロジスティック回帰は、機械学習におけるクラスを分類するために一般的に使用される方法であり、「一般線形モデル」に属する。そのうちに、 $y \in \{0,1\}$ 。最も理想的なのは「単位ステップ関数」である。すなわち、予測値が0より大きい場合は正の例であると判断し、0より小さい場合は負の例と判断し、臨界値が0の場合は任意に決定することができる。

5-4-3 決定木

いくつかの属性(特徴)により、各ノードは質問をする。判断することによって、データは2つのクラスに分けられ、また質問される。これらの問題は、既存のデータから学習される。新しいデータが追加されると、決定木上の問題に応じてデータを適切な葉に分割することができる。決定木は木構造をしているので、ルートノードは唯一である。決定木により、データから最も良い決定木を構築できる。この過程は決定木学習と呼ぶ。決定木は一定程度にデータ構造学科の二分木と似ている。

ルート：学習の例の集合。

分岐：分類の基準。

リーフ：各クラス。

例の図解は以下に示す。

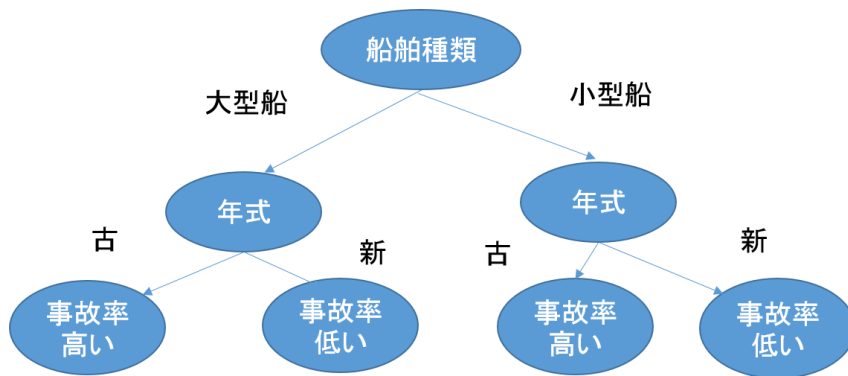


図 3.決定木の図解

決定木構築の基準として、最も小さい決定木を構築する。

5-4-4 ランダムフォレスト

ランダムフォレストは決定木の融合である。ランダムフォレストは行列を見ると、決定木は子行列に相当する。まず、ランダムフォレストから N 個の子行列を生成する。その後、 N 個の子行列から N 個の決定木を生成し、 N 個の決定木からの一番良い決定木の分類結果を最後の予測結果とする。

5-4-5 KNN(k-nearest neighbor classification)

KNN アルゴリズム[22]は、 k -最近傍クラス分類アルゴリズムとも呼ばれる。KNN アルゴリズムは、訓練集合から新しいデータに最も近い k 個のデータを見つけ、次いで、それらの主な分類に基づいて新しいデータのカテゴリを決定する。このアルゴリズムは、学習データセット、距離または類似程度、および k の大きさの 3 つの主要因を含む。

アルゴリズムの手順は次のとおりである。

- 1) 距離の計算：テストデータが与えられたら、学習データセット内の各データからの距離(類似程度)を計算する
- 2) 近傍を見つける：最も近い k 個の学習データをテストデータの最近傍として囲む
- 3) 分類： k 個の近傍の主なカテゴリに基づき、テストデータが分類される。

5-4-6 k-means

K-means は、クラスタリングの問題を解決する教師なしアルゴリズムである。所与のデータセットは、ある数のクラスタ (k 個のクラスタを仮定) によって分類される。アルゴリズムの手順は次のとおりである。

- 1) K-means は各クラスタの k 個の点を選択し、重心と呼ばれる。
- 2) 各クラスタ点は、最も近い重心 (k 個のクラスタ) を有するクラスタを形成する。
- 3) 既存のクラスタメンバーに基づいて各クラスタの重心を見つける。ここに新しい

重心を取得できる。

- 4) 新しい重心があるので、手順2と3を繰り返す。新しい重心から各データ点に最も近い距離を見つけ、そのクラスターへ割り当てを変える。このプロセスは、収束が生じるまで繰り返される。すなわち、重心は変化しない。

5-4-7 Nave Bayes[23]

Nave Bayes は確率に基づいて予測するアルゴリズムである。前の節のロジスティック回帰は、フィッティング曲線による分類を達成する。決定木は、最適な分類機能を見つけてサンプル経路を学習することによって分類を学習する。対照的に、Nave Bayes は、特徴確率を考慮して分類を予測する独特の手法である。

ベイズの定理は、ランダム事件 A および B の条件付き確率に関する定理である。そして、ベイズの式は以下に示す。

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (5-4-2)$$

P(A):A 事件の発生確率

P(B):B 事件の発生確率

P(A|B):A 事件に対して B 事件の発生確率

P(B|A):B 事件に対して A 事件の発生確率

アルゴリズムの手順は次のとおりである。

- 1) データの特徴属性 A_i を確認する上、学習データを獲得する。
- 2) クラス別に $P(B_i)$ を計算する。
- 3) 特徴属性ごとに $P(A_i|B_i)$ を計算する。
- 4) クラス別に $P(A_i|B_i) * P(B_i)$ を計算する。
- 5) $P(A_i|B_i) * P(B_i)$ の極大項を A_i の所属クラスとする。

5-4-8 SVM

SVM(Support Vector Machine)は1963にVladimir N. Vapnikら[24]から提案されたパターン認識に関する教師あり学習モデルである。同時に、サポートベクターマシンは分類と回帰の属性を持たずアルゴリズムである。また、SVMは2分類のアルゴリズムに属し、線形と非線形分類をサポートする。現在、SVMは複数の分類をサポートできるが、2分類でより多く応用されている。

5-4-9 ニューラルネットワーク

ニューラルネットワーク(Neural Network, NN)は人間の脳の神経回路に基づいて提案され、歴史上多くの異なるバージョンがある。ニューラルネットワークは人間の脳神経と似ているから、人工ニューラルネットワークとも呼ばれる。ニューラルネットワークの構造は以下の図4.の示すようである。

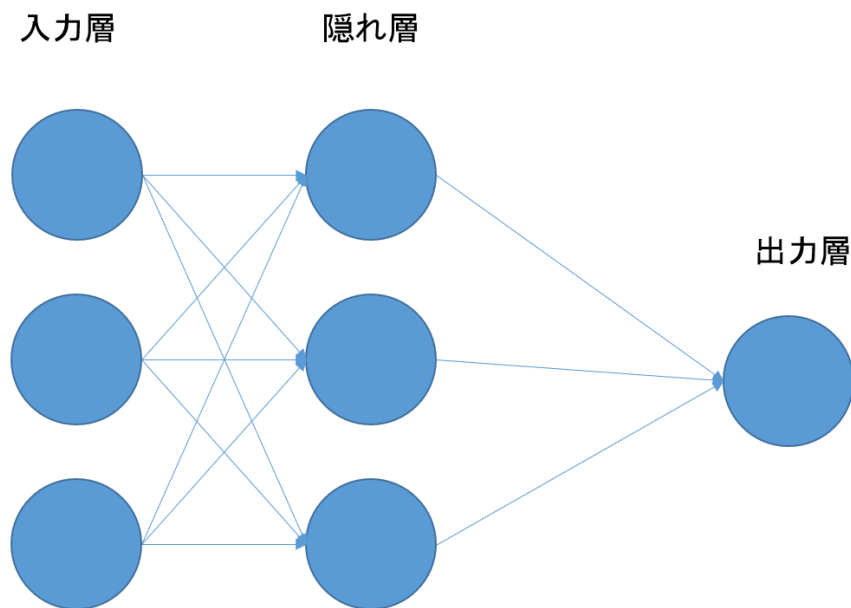


図4. ニューラルネットワークの構造

通常、ニューラルネットワークは、入力層、複数の隠れ層、および出力層で構成されている。図4中の円は、ニューロン（パーセプトロンとも呼ばれる）として考えることができる。ニューラルネットワークでは重要な作業は、隠れ層とニューロン間の重み（weight）を計算することである。上図の示すように隠れ層が少ないニューラルネットワークはSNN（shallow neural network）と呼ぶ。隠れ層が多層であるニューラルネットワークはDNN（deep neural network）である。ニューラルネットワークは分類問題において良好な機能を持たす。上述の線形回帰または線形SVMは線形分類に適用するが、非線形分類に良い効果がない。ニューラルネットワークの計算はより簡単である。ニューラルネットワークの最も有名なアルゴリズムは誤差逆伝播（Error Back Propagation, BP）である。ニューラルネットワークの用途はn個入力データが一系列の計算を通じ、m個確率を出力する。

アルゴリズムの手順は次のとおりである。

- 1) 入力データが線形変換 $wx + b$ で処理し、線形データへ変更する。
- 2) 線形変換の後、線形データが活性化関数をかけて非線形データを得る。活性化関数はさまざまな種類があり、例えば、sigmoid, tanh や relu などの関数がある。活性化関数の目的は線形結果を非線形化する。活性化関数を追加しないと、何層の隠れ層があっても、最終の出力は元の入力された線形変換である。そのため、多層隠れ層の存在は意味がない。そして、各隠れ層には、非線形変化をもたらす活性化関数が備わっている。

- 3) 1) 2) ステップは学習の損失 (loss) を求める。最後、誤差逆伝播を利用して隠れ層と出力層の誤差を取得し、その誤差により各層の重みを修正する。上記のプロセスを通じ、全ての重みを更新することができ、条件が満たされるまで繰り返す。

6、深層学習による物体認識

6-1 深層学習

2010 年ごろになると、様々な改良を経てきたニューラルネットワークの技術が深層学習（またはディープラーニング）と呼ばれる[25]。深層学習は機械学習の一種類で、人間の脳神経構造みたいニューラルネットワークを用いた機械学習である。一方、深層学習は人工神経ネットワークとも呼ばれ、多層ニューラルネットワークを用いて学習能力を構築するネットワークである。深層学習の概念は最初 2006 年に Hinton らから提出した。深層信念ネットワーク（DBNs, Deep Belief Nets）に基づいて、深層構造に関連する最適化問題を解決することを望むために、各層の学習アルゴリズムが提案され、多層自動エンコーダの深い構造が提案される。さらに、Lecun らによって提案された畳み込みニューラルネットワークは、最初の本格の多層構造学習アルゴリズムである。このアルゴリズムは訓練性能を改善するために空間的相対関係を用いてパラメータ数を減らした。

深層学習は近 20 年以來重大な発展があり、様々な分野で活用している。たとえば、データ抽出、コンピュータ視覚、自然言語処理、音声認識などの分野、特に画像認識と音声認識の分野で広く応用されている。機械学習と同じ、深層学習も 3 種類に分けられ、教師あり学習、教師なし学習、強化学習がある。そのうちに、教師あり学習に基づく深層学習モデルは畳み込みニューラルネットワーク（CNNs, Convolutional neural networks）である。深層信念ネットワーク（DBNs, Deep Belief Nets）は教師なし学習に基づく学習モデルの一種類である。

6-2 畳み込みニューラルネットワーク

深層学習に最も広く応用されている分野の 1 つは、深い畳み込みニューラルネットワーク（CNNs, Convolutional neural networks）である。畳み込みニューラルネットワークは人間の脳みたいニューラルネットワークである。たくさんの神経元があり、それぞれの神経元は一つの数値を持っている。神経元と神経元の間には特殊な計算あるいは関数を通じ、最後予想の出力値を出力する。畳み込みニューラルネットワークは、近年発展した有効な物体識別方法であり、広く注目されている。1960 年代に、Hubel と Wiesel がネコの大脳皮質のニューロンを研究したとき、彼らは独自のネットワーク構造がフィードバックニューラルネットワークの複雑さを効果的に減らし、次に畳み込みニューラルネットワークを提案した。現在、CNNs は多くの科学分野、特にパターン分類の分野で応用されている。また、CNNs は画像の複雑な事前処理がいらないため、原画像を直接に入力することができるという利点がある。

6-3 今まで流行している画像認識手法(CNN に基づく)

- CNN(Convolutional neural networks) [27]
- R-CNN(Regions with CNN features) [28]
- Fast R-CNN[29]
- Faster R-CNN[30]
- SPPnet (Spatial pyramid pooling in deep convolutional networks) [31]
- SSD (Single Shot MultiBox Detector)[32]
- YOLO(You Only Look Once)[33]

以上の画像認識アルゴリズムは全部 CNN に基づく学習モデルである。そのうちに Faster RCNN,YOLO,SSD この三つはよく使われているアルゴリズムである。本研究では SSD(tensorflow-keras)を利用して、船舶の認識研究を行った。

6-4 深層学習フレームワーク

深層学習フレームワークは、深層学習のライブラリーである。プログラミングの際にインポートする必要がある。深層学習フレームワークの中に、いろいろな深層学習関数があり、インポートすれば直接に使用できる。現在、様々な深層学習フレームワークが開発され、各深層学習フレームワークは自分の特性を持っている。現在流行している深層学習フレームワークは主に 6 種類があり、それぞれは Caffe、TensorFlow、Torch、Theano、MXNet、Kares である。この節では、本研究で使ったフレームワーク Caffe,TensorFlow,Kares を紹介する。

Caffe は最も古いフレームの 1 つで、非常に良い機能を持っている。同時に、汎用性を持っている。Caffe フレームワークの学習時間は、Keras-Theano より 5 倍ほどは早い。Caffe 1.0 は、最初に深い学習ツールとして主流となっている。Caffe は C および C ++ に基づく深層学習フレームワークで、特に畳み込みニューラルネットワークの実装上に非常に優しい特性がある。今まで、Caffe はコンピュータ視覚の分野で最も人気のあるツールキットとして存在している。しかし、初期の開発と歴史的な問題のために、十分に柔軟ではなく、再帰的なネットワーク RNN (Recurrent neural network) [26]と言語モデリングのサポートが不足であるという欠点がある。そのため、Caffe はテキスト、音声、時系列データなどの深層学習には適用しない。しかし、非常に安定し、高速なフレームワークである。

TensorFlow は、DistBelief に基づく Google の第 2 世代の人工知能学習フレームワークであり、Tensor は N 次元配列を意味し、Flow はデータフローグラフに基づく計算を意味する。TensorFlow は、テンソルがフローグラフの一端から他端に流れる計算プロセスである。分析と処理のために複雑なデータ構造を人工知能ニューラルネットワークに入力フレームワークである。また、TensorFlow は、音声認識や画像認識などの多くの機械学習分野で使用することができる。2011 年に開発された深層学習基礎フレームワーク DeepBelief の各方面を改善した。今 TensorFlow は完全にオープンソースで誰にでも利用可能である。

TensorFlow は Python で書かれ、C/C++エンジンによって加速でき、Google のオープンソースの第二世代の深い学習フレームワークである。TensorFlow は再帰ニューラルネットワーク RNN を非常に扱いやすく、内部実装ではベクトル演算ができ、新しいネットワークの設計を非常に容易にし、迅速な開発をサポートするフレームワークである。TensorFlow の応用は深層学習だけではなく、強化学習やその他のアルゴリズムにもサポートしているので、十分な柔軟性を持っている。欠点として、TensorFlow は計算量が膨大なため、システム実行の効率が低下し、メモリの使用量が多い。

現在流行しているフレームワークはもう一つがある。それは大量の計算過程を隠れたフレームワーク keras である。Keras は二つ backend tensorflow と theano があり、すごく使いやすい。また、keras ではたくさんの関数が直接に呼び出しでき、拡張性を持っている。さらに、keras ライブラリーのソースコードは python に基づくため、読みやすい。そのため、本研究では SSD が keras に基づき、tensorflow が backend とする。

6-5 深層学習手法

● CNN

一般物体認識の主要なタスク

- (a) 物体カテゴリ識別
- (b) 物体検出
- (c) 物体領域抽出と表示

CNN は、畳み込み層とプーリング層と呼ばれる 2 種類の層を持っているフィードバック型の NN である。CNN の例として、簡単な CNN ネットワークの構造図は以下に示す。

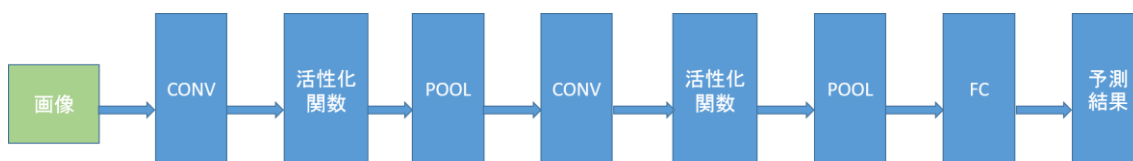


図5. CNN ネットワークの構造

畳み込みニューラルネットワーク (CNN) は、入力層、畳み込み層、活性化関数、プーリング層、および全結合層から構成する。

- (1) Convolution Layer (畳み込み層)：特徴量の畳み込みを行う層である。
この層を通じ、画像の特徴を抽出できる。一般的に、フィルタを用いて元画像とかけ、画像の feature map を獲得する。
- (2) 活性化関数：ニューラルネットワークにおいて、活性化関数の役割は、ニューラルネットワークにいくつかの非線形因子を加えることであり、その結果、ニューラルネットワークはより複雑な問題をよりよく解決することができる。活性化関数は主に四種類があり、それぞれは sigmoid, tanh, ReLU, softmax 関数であ

る。CNN には一般に ReLU 活性化関数がよく使われている。

1) sigmoid 関数

隠れ層ニューロンの出力にはよく使われている。

$$s(x) = \frac{1}{1+e^{-x}} \quad (6-5-1)$$

この関数は出力が $[0, 1]$ の間にあり、出力層に使用され、関数の微分を求めることが容易である。しかし、勾配消失の問題があるため、学習の時にトレーニング結果に影響を与える。

2) tanh 関数

$$\tanh(x) = \frac{1-e^{-2x}}{1+e^{-2x}} \quad (6-5-2)$$

tanh 関数の値域は $[-1, 1]$ である。この関数はシグモイド関数よりも速い収束速度という利点を持つが、シグモイド関数と同じ欠点、勾配消失の問題がある。

3) ReLU (Rectified Linear Unit Function) 関数

シグモイド関数と同じ、隠れ層ニューロンの出力に使われている。

$$f(x) = \max(0, x) \quad (6-5-3)$$

ReLU 関数は $x \geq 0$ の時に、 $f(x)=x$ 。その以外の場合には、 $f(x)=0$ 。

利点：ReLU 関数は線形関数であるため、sigmoid と tanh 関数に比べて迅速に収束することができる。さらに、勾配消失の問題を効果的に緩和する。

欠点：学習が進むにつれて、ニューロンが学習できない場合もあるので、ニューロンの間の重みが更新できないことがある。

4) Softmax 関数

多種類ニューラルネットワークの出力に使われている。

$$f(x) = \log(1 + \exp(x)) \quad (6-5-4)$$

Softmax 関数は、複数のニューロンの出力を $(0, 1)$ 値域にマッピングする。

また、この値域に基づいて分類する。この過程は複数の分類に使用される確率として理解できる。

- (3) Pooling Layer (プーリング層)：畳み込み層からの feature map の縮小を行い、扱いやすくするための層である。換言すればフィルタを用いた特徴が鮮明になるための処理層である。プーリング層を通じ、ある程度上に計算量を削減した。プーリング層には、四つの主な機能がある。次元の削減、非線形の実現、受容野の拡大、不変性を達成することができる。不変性は変換不変性、回転不変性、スケール不変性を含む。最も使っているプーリング方法は二つがあり、平均プーリングと最大プーリングである。

1) 平均プーリング

画像領域の平均を計算して、この領域のプーリングされた値とする。

$i*j$ の大きさの受容野 p_{ij} と仮定すると、平均プーリングでプーリングされた値は

$$y_{ijk} = \frac{1}{|p_{ij}|} \sum_{(p,q) \in p_{ij}} x_{pqk} \quad (6-5-5)$$

2) 最大プーリング

画像領域の最大値を見つけ、この領域のプーリングされた値とする。

平均プーリングでプーリングされた値は

$$y_{ijk} = \max_{(p,q) \in p_{ij}} x_{pqk} \quad (6-5-6)$$

(4) Fully Connected Layer(全結合層):特徴量から、最終的な判定を行う層である。

全結合ネットワーク層は各ノードが前の層のすべてのノードに接続され、前の処理から抽出されたフィーチャを結合する。完全に結合されているため、全結合ネットワーク層のパラメータは最も多い。

● R-CNN

R-CNN (Regions with CNN features) は CNN 手法に基づいて、物体検出問題に適用するマイルストーンであり、2013 に Ross Girshick らから提案されている。CNN の優れた特徴抽出と分類性能により、候補領域生成 (Region Proposal) 手法によって物体検出と領域抽出を実現した。

RCNN のネットワーク構造は以下に示す。

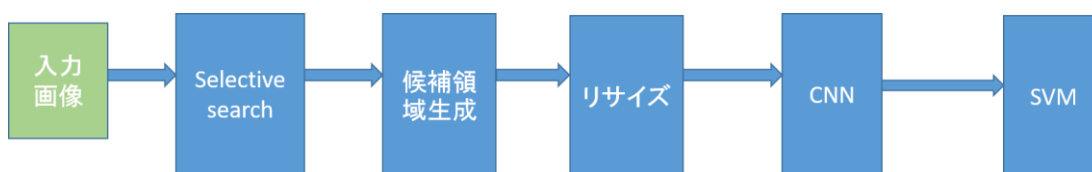


図 6. RCNN ネットワークの構造

- (1) 画像を読み込む。
- (2) Selective Search [34] (似た領域を階層的にグルーピングし、物体領域の候補を出す) で候補領域 (約 2000 個程度) を出す。
- (3) 候補領域をそれぞれリサイズし、CNN に入力する。
- (4) CNN からの特徴量を SVM 分類器で分類し、矩形の座標を活性化関数で回帰する (候補領域のズレを補正する)。

- Fast R-CNN

Fast R-CNN この学習モデルは、R-CNN の 3 つの問題を解決した。R-CNN では、画像内の多数の候補領域が重なり、特徴を抽出する計算は冗長性がある。この論文では、画像全体が正規化され、深層ネットワークに直接入力される。接続する場合、候補領域の情報が追加され、各候補領域は最後のいくつかの層で処理される。R-CNN の学習速度が遅い問題に対して、この論文では最初に画像をネットワークに入力し、その後に画像から抽出した候補領域を入力する。これらの候補領域の最初のいくつかの特徴は、繰り返す計算の必要はない。R-CNN 中の独立分類器と回帰関数は、学習サンプルとして大量な特徴が必要なため、学習に必要なスペースが大きい。この問題に対して、Fast R-CNN はカテゴリ判定と位置の補正を深層ネットワークで統一に実現し、追加のストレージがいらない。

Fast R-CNN のネットワークの構造は以下に示す。

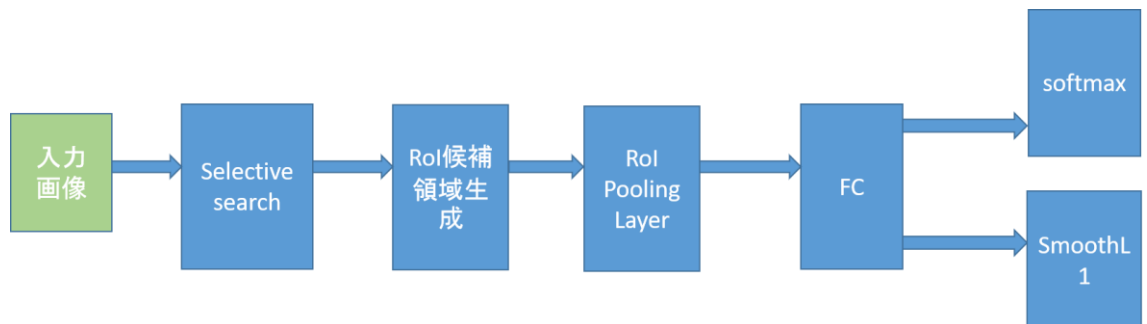


図 7. Fast RCNN のネットワークの構造

- (1) 画像を読み込む。
- (2) Selective Search で候補領域 RoI (Regions of Images) を出す。
- (3) 候補領域をそれぞれリサイズし、CNN に入力する
- (4) CNN からの feature map を RoI pooling layer に入力する。RoI Pooling で任意サイズの領域をプーリングして、固定サイズの出力をする。普通の Max Pooling (フィルタ出力層の小領域をまとめる処理)と同様に誤差に誤差逆伝播ができる。
- (5) SoftmaxLoss は SVM を置き換え、SmoothL1Loss は Bouding box の回帰を置き換える。

- Faster R-CNN

Faster R-CNN の構造

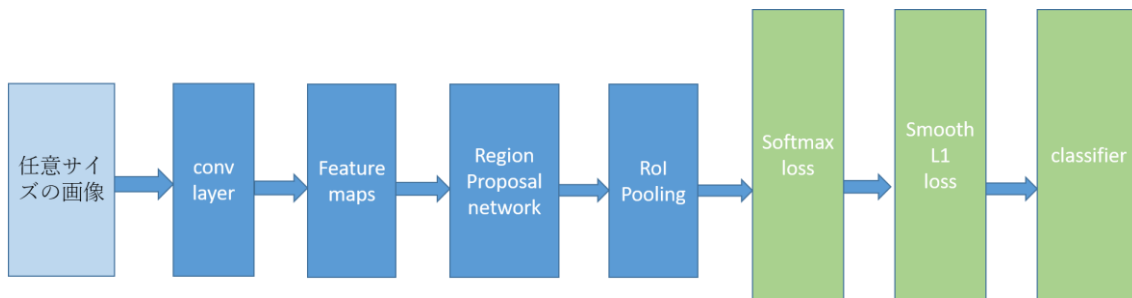


図 8. Faster RCNN の構造

- (1) 任意サイズの画像を入力する。
- (2) 特徴抽出のために画像全体を CNN に入力する。
- (3) CNN から取得した feature map を RPN (region proposal network) に入力する。
- (4) RPN で (region proposal) を生成し、各画像に対して 300 個の候補領域を生成する。
- (5) 候補領域を CNN の畳み込み特徴マップの最後のレイヤにマッピングする。
- (6) RoI プーリング層を通じ、各 RoI に固定サイズの特徴マップを生成する。
- (7) 分類訓練とボックス回帰に Softmax Loss (分類確率検出) と Smooth L1 Loss (ボックス回帰検出) を使用する。

RPN ネットワークの構造

Region Proposal Network(RPN)領域生成ネットワークというのは物体候補領域を推定してくれるネットワーク、つまり任意サイズの画像を入力して、物体候補領域とそのスコアを出力するネットワークである。RPN は走査窓を利用して畳み込み層からの feature map を走査する。各走査窓の中心には、3 種類スケール{128,256,512}と 3 種類のアス比{1 : 1, 1 : 2,2 : 1}のアンカー (候補ウィンドウ) で走査する。また、トレーニングセットの中の各画像を考察する。各較正された真値候補領域 (proposal) について、その重なる割合(IoU Intersection-over-union)が最も高い Anchor が真値サンプルとして記録される。残りの Anchor には、IoU が 0.7 より大きい場合、合格サンプル(positive)として記録され、IoU が 0.3 未満である場合、不合格サンプル(negative)として記録される。それ以外の Anchor は、捨てる。さらに、画像のエッジを越えて破棄される。

IOU 閾値：物体検出では、物体の bounding box を定位する必要がある。アルゴリズムからのデータが手作業でラベル付けされたデータと完全に一致することができないため、位置決め精度評価式である。IOU は、2 つの bounding box の重なる程度を定義する。たとえば、手作業で付けされたボックス A と Selective Search で付けされたボックス B の一致度

IOU の公式は

$$\text{IOU} = \frac{S_1}{(S_A + S_B - S_1)} \quad (6-5-7)$$

S_1 : 重なるの面積

S_A : ボックス A の面積

S_B : ボックス B の面積

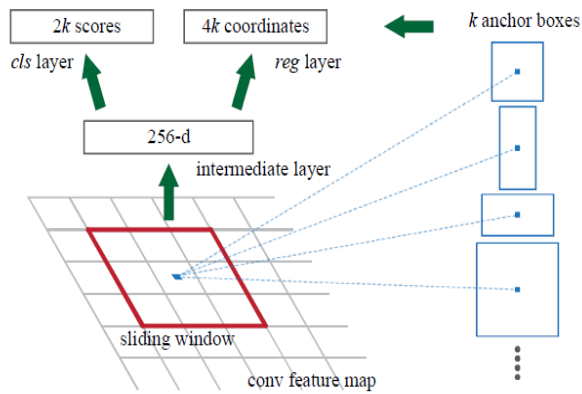


図 9. RPN の構造[31]

FASTER-RCNN は、畳み込みネットワークを創造的に使用して候補領域を生成し、畳み込みネットワークと目標検出ネットワークが共有するので、候補領域数が約 2,000 から 300 に減少し、候補領域の品質も改善される。

- SSD

- SSD の構造

- 本研究使っている SSD のフレームワークは tensorflow-karas である。

- SSD は多種の異なるサイズの feature map を利用したため、物体が小さくても検出できる。

- SSD の学習モデルは以下の図に示す。

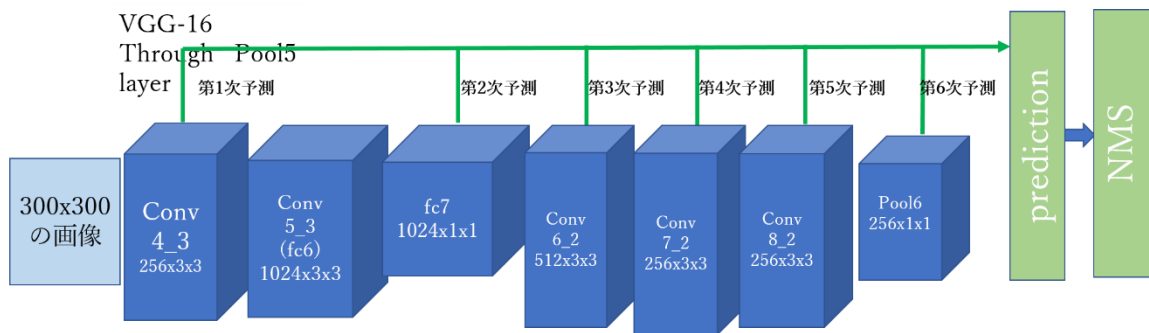


図 10. SSD の構造

- (1) 300*300 の画像を読み込む。元画像に $M \times N \times 3$ のフィルタをかけ、特徴を抽出する。「300x300x3 の画像」は 300x300 ピクセルの画像、3 は色情報 (R、G、B) である。
- (2) VGG16 のネットワークを通じ、 $256 * 3 * 3$ チャンネルの conv4_3 から予測する。
- (3) $1024 * 3 * 3$ の conv5_3 から第一次 FC 処理を行った後、もう一度 FC 処理を行い、第二次予測する。
- (4) $512 * 3 * 3$ の conv6_2 から第三次予測を行う。
- (5) $256 * 3 * 3$ の conv7_2 から第四次予測を行う。
- (6) $256 * 3 * 3$ の conv8_2 から第五次予測を行う。
- (7) $256 * 1 * 1$ の pool6 から第六次予測を行う。
- (8) 最後、NMS 関数 (非極大値抑制関数, Non-maximum suppression) で一番高いスコアのボックスを出力する。NMS 関数は最も信頼度の高いものが最終結果として選択する。重なりがない場合は、最終結果としてそのまま保持する。二つ以上出力される場合、認識率の高いボックスのみ出力する。以下の図 11 は一つの例である。図 11 では 100% 確率のボックスと 97% 確率のボックスがあるが、100% 確率のボックスのみを表示される。97% 確率のボックスは捨てる。

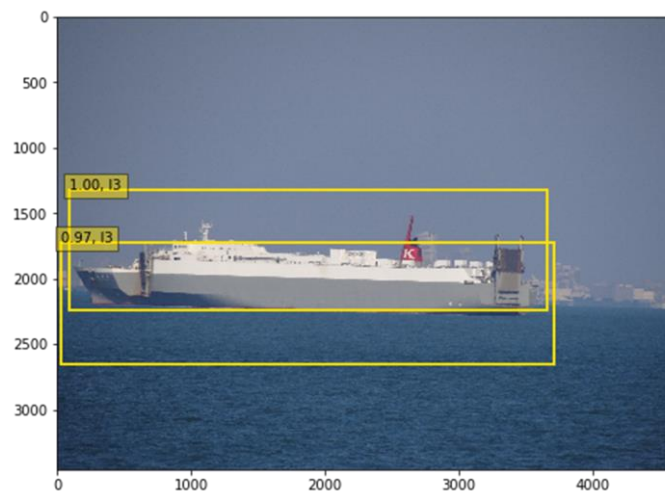
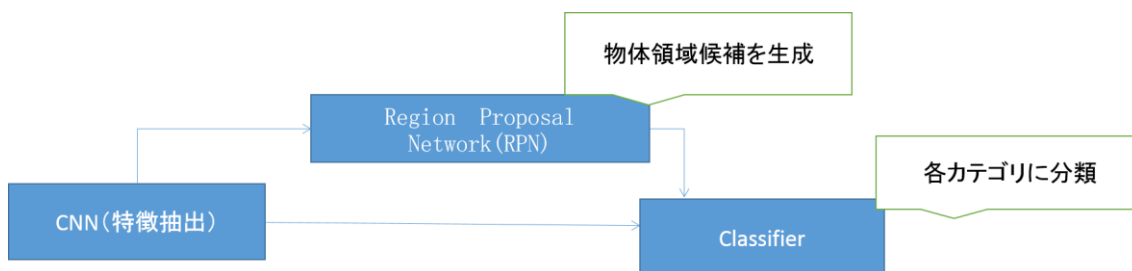


図 11.NMS 関数の例

Faster R-CNN と SSD の比較

Faster R-CNN



SSD



図 12. Faster RCNN と SSD の比較

SSD の検索速度と検索精度は Faster R-CNN より高い。また、SSD は比較的 low 解像度 (300 x 300px) でも高精度に検索できる。階層的な Feature map と利用することでさまざまなスケールに対応する。アスペクト比ごとに識別器を作ることで、高い精度の検出率を達成した。SSD はシンプルなネットワーク設計で、Faster R-CNN より簡単である。

7、航行中の船舶からの物体の認識

7-1 データセット 1

船舶認識実験は二つのデータセットを作った。2017年9月21日、22日及び10月6日に東京海洋大学所有船舶「汐路丸」(全長:49.63[m]、幅:10.0[m])を使用し、学習に使用する海上環境画像を撮影した。取得した約4000枚の画像はデータセット1を作成し、9カテゴリを分類し、約50epochを学習した。障害物からコンテナ船、LNG (liquefied natural gas)船、PCC (自動車運搬船)、その他の大型船、漁船、プレジャーボート、タグボート、その他の小型船、ブイの9つのカテゴリに分類する。データセット1の例は以下に示す。

コンテナ船(11)

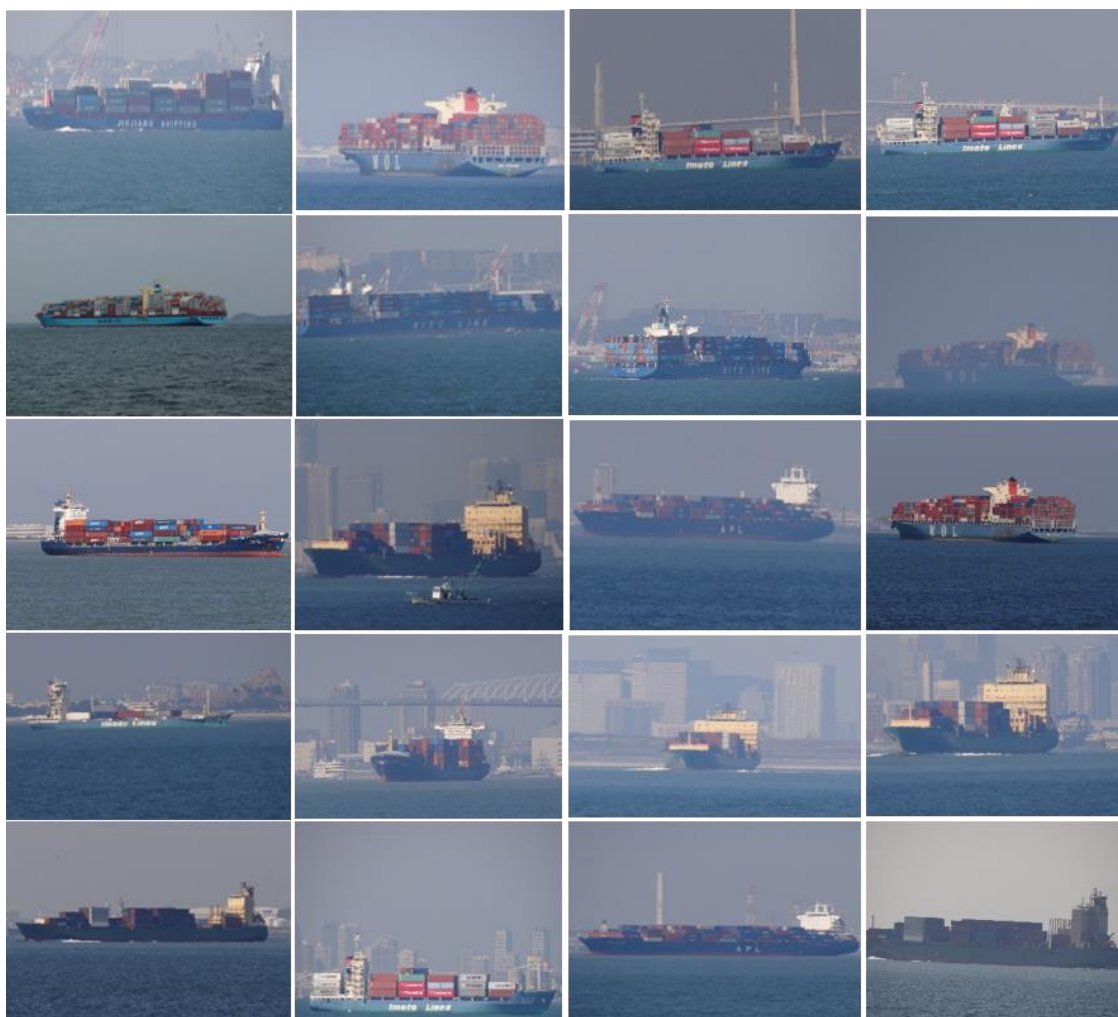


図 13.コンテナ船の画像

LNG 船(12)

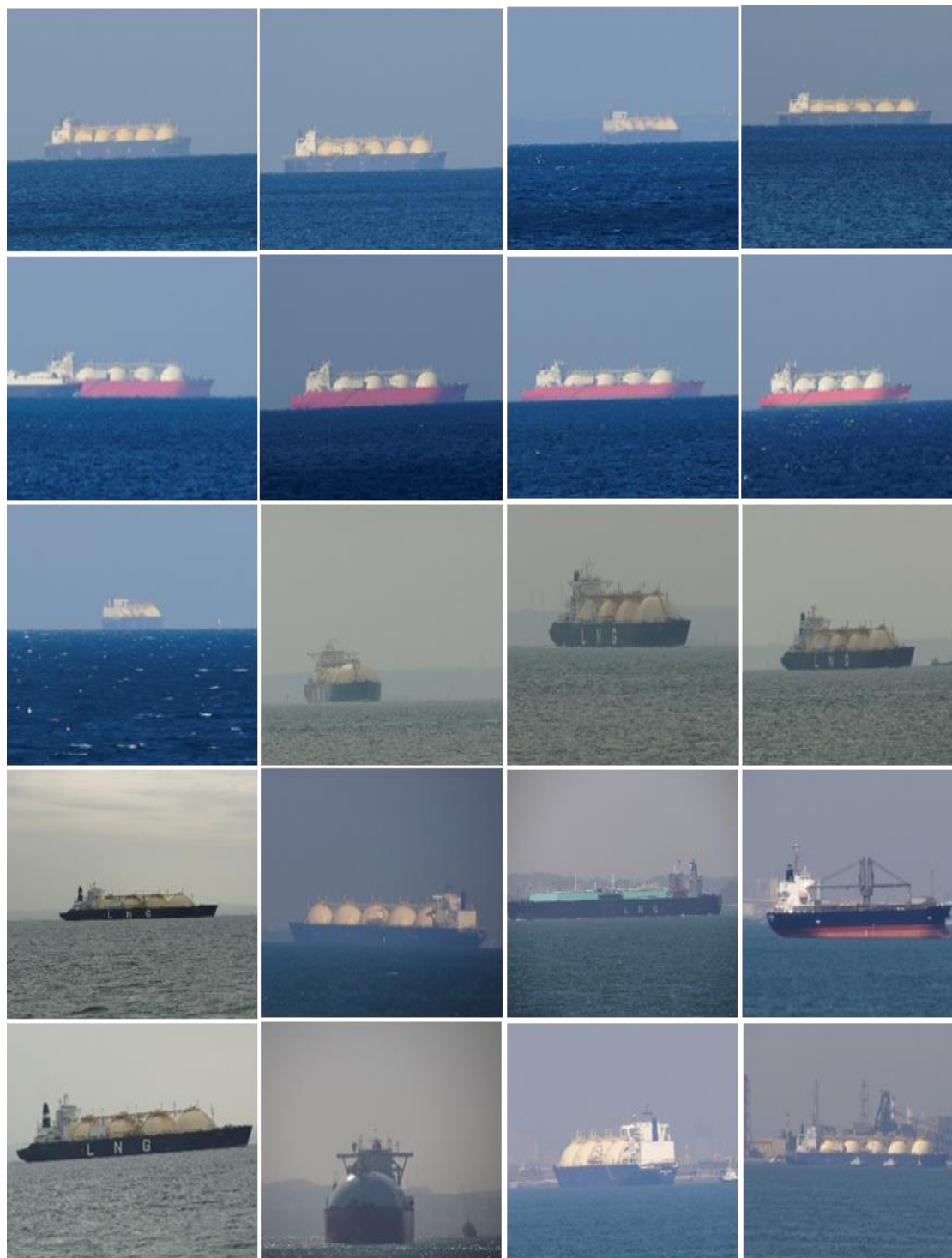


図 14.LNG船の画像

PCC 船(13)

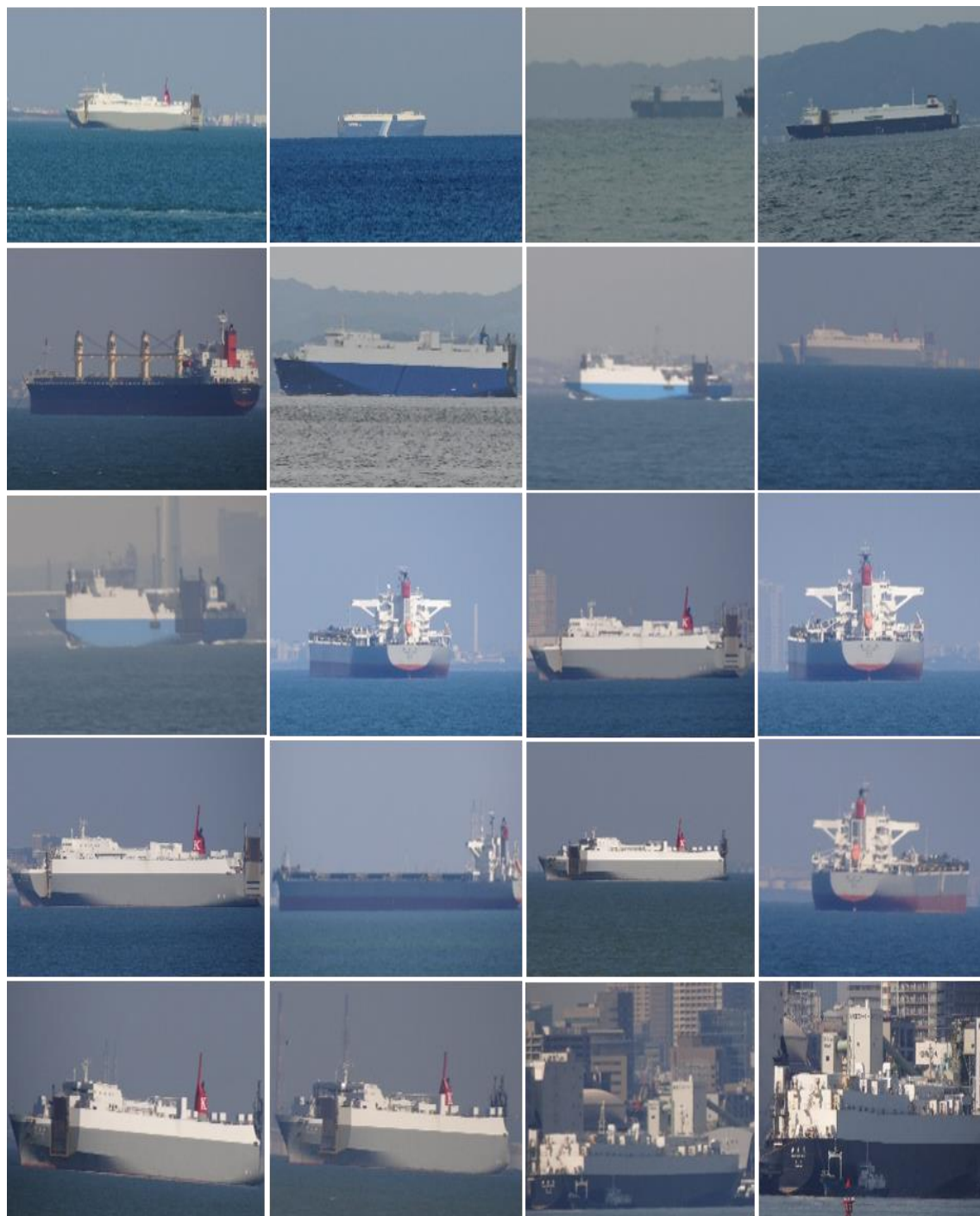


図 15. PCC 船の画像

他の大型船 (14)

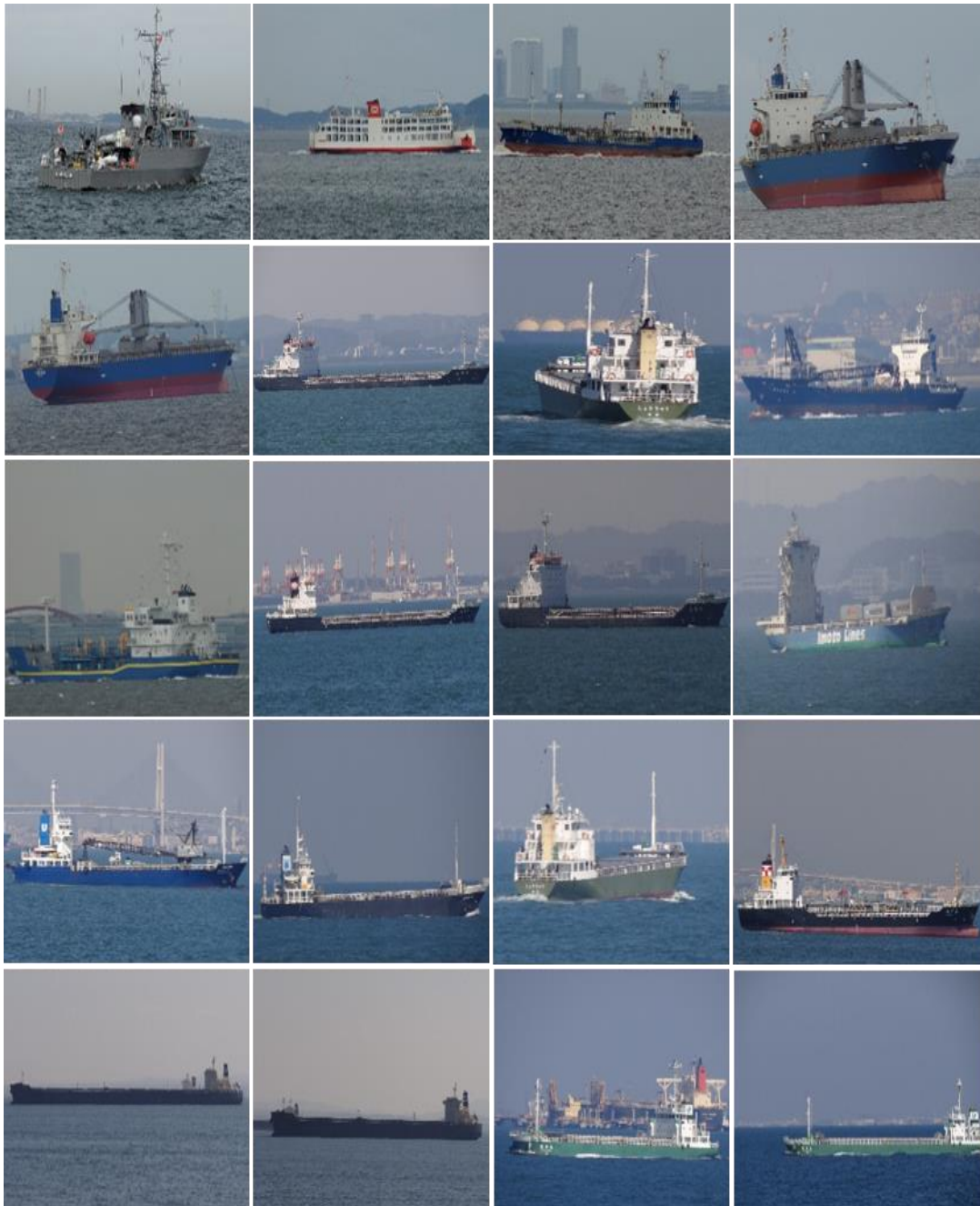


図 16.他の大型船の画像

漁船 (s 1)



図 17. 漁船の画像

プレジャーボート (s2)

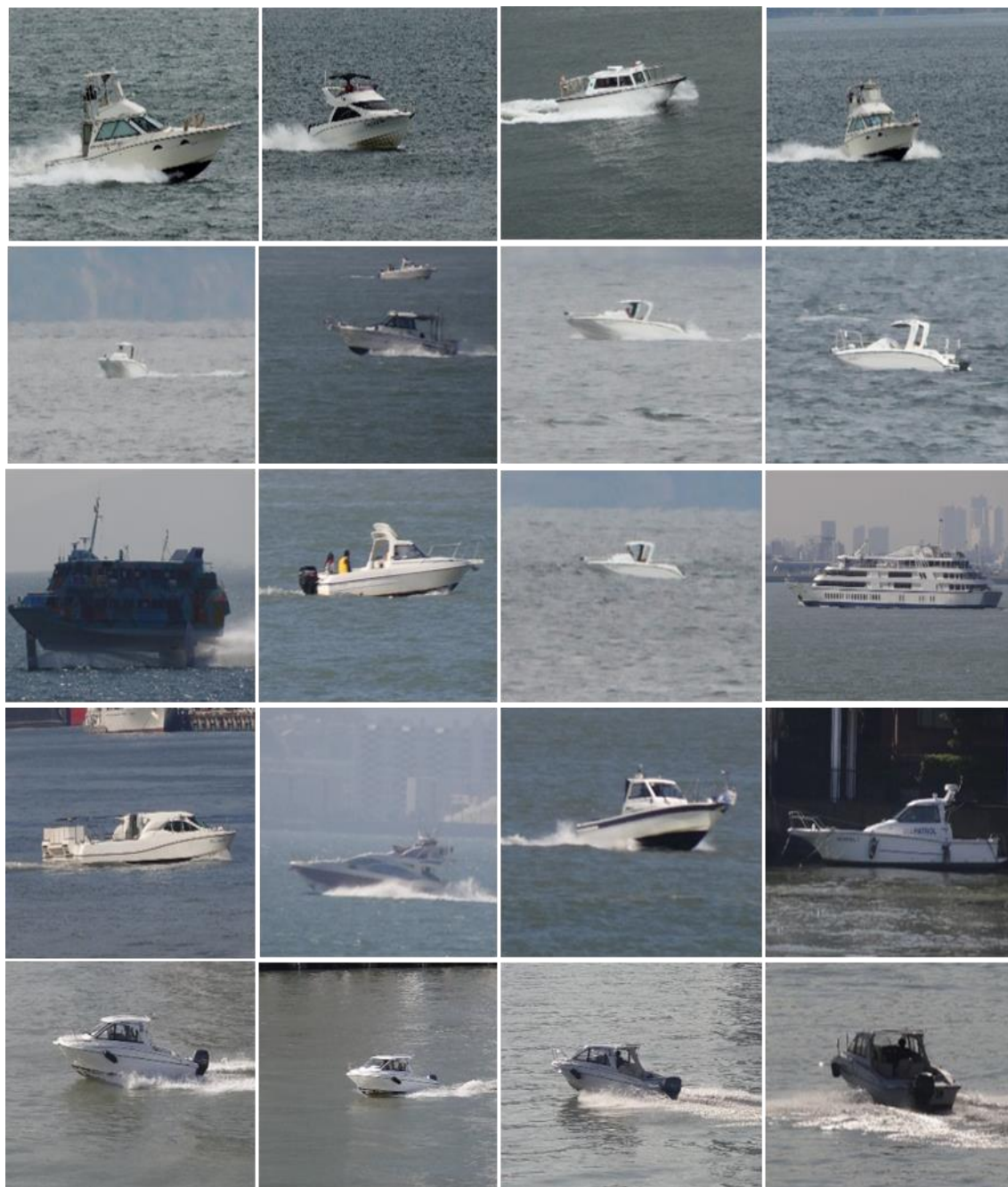


図 18. プレジャーボートの画像

タグボート (s3)

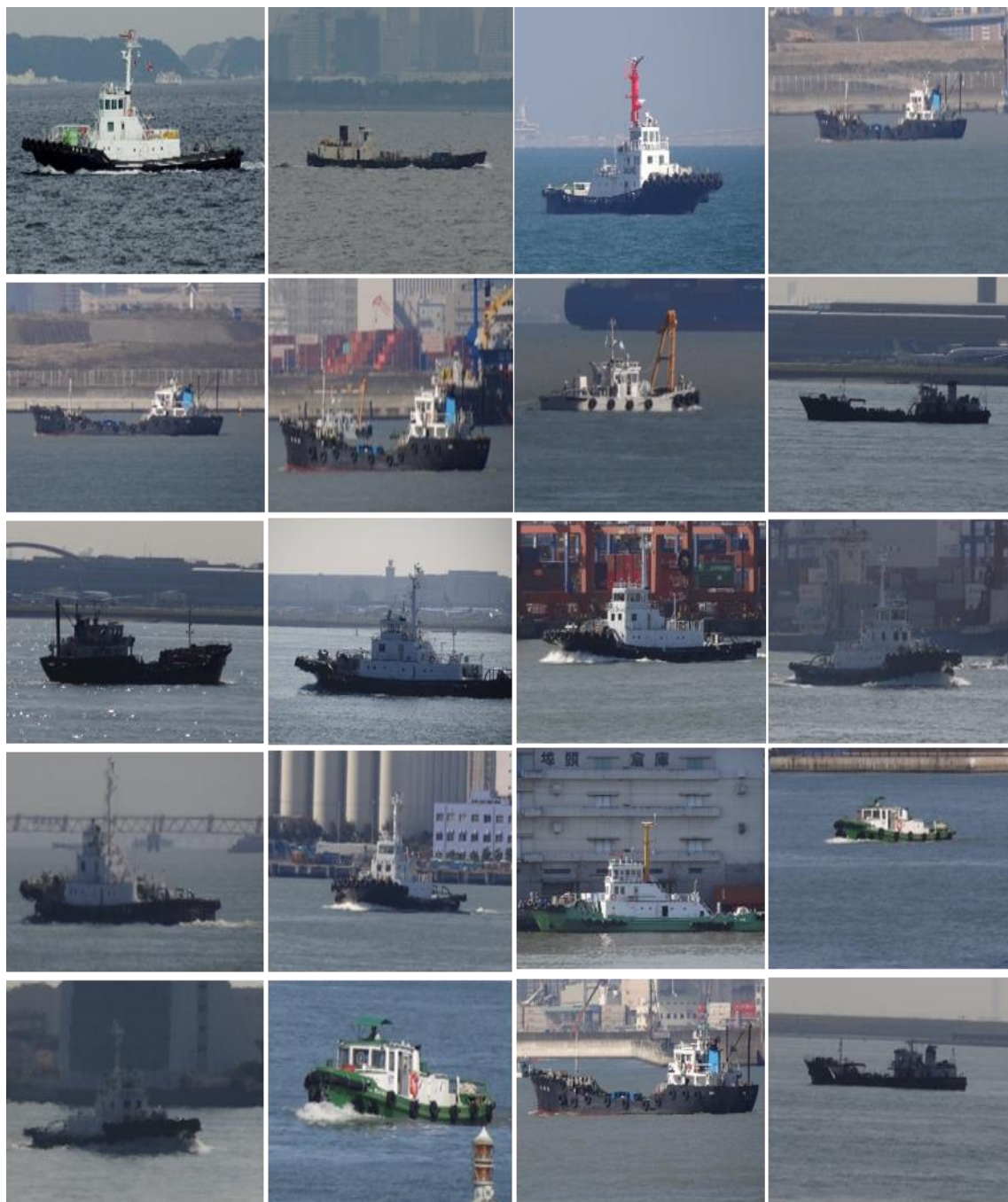


図 19. タグボートの画像

他の小型船（s 4）

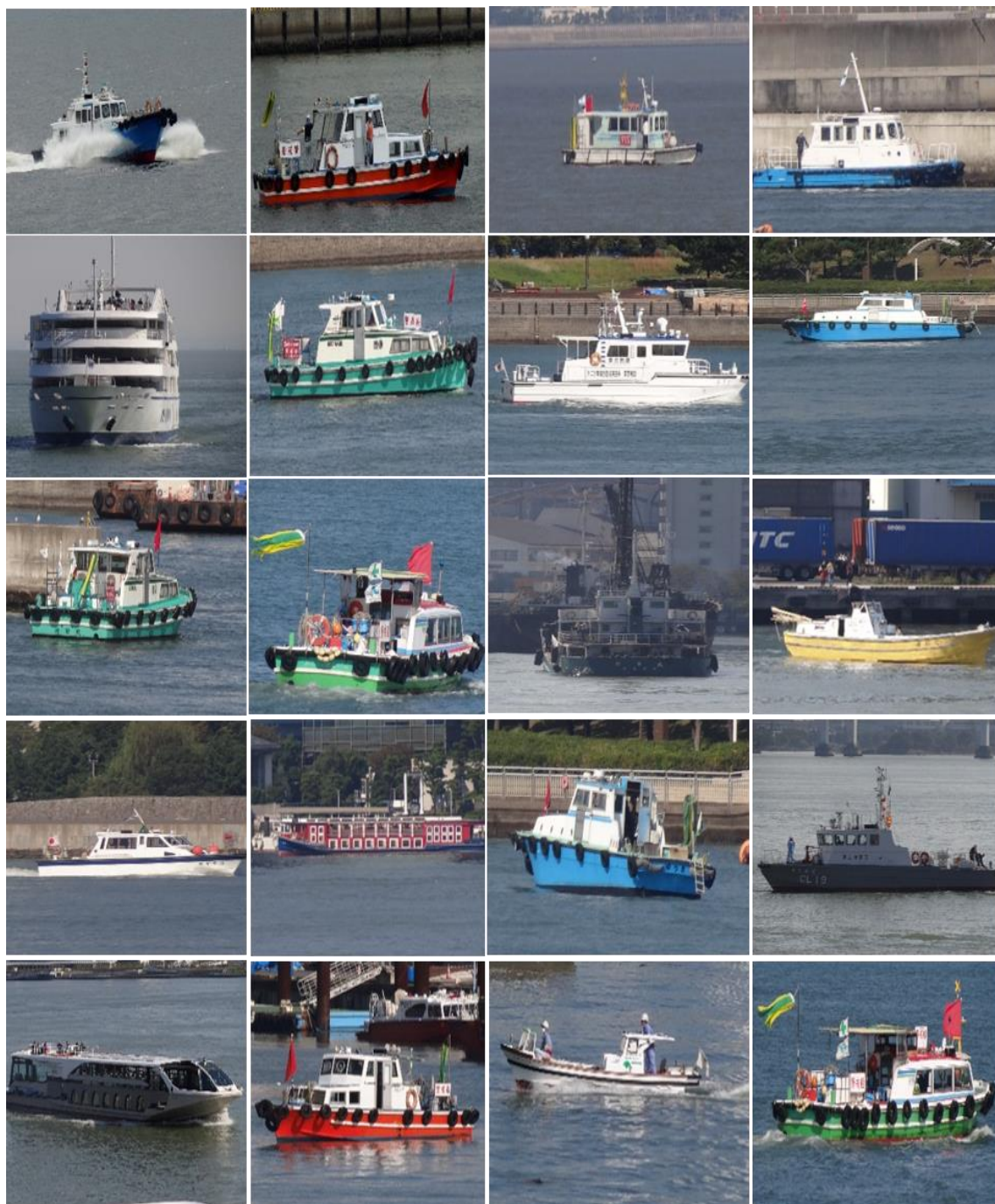


図 20.他の小型船の画像

ブイ (buoy)

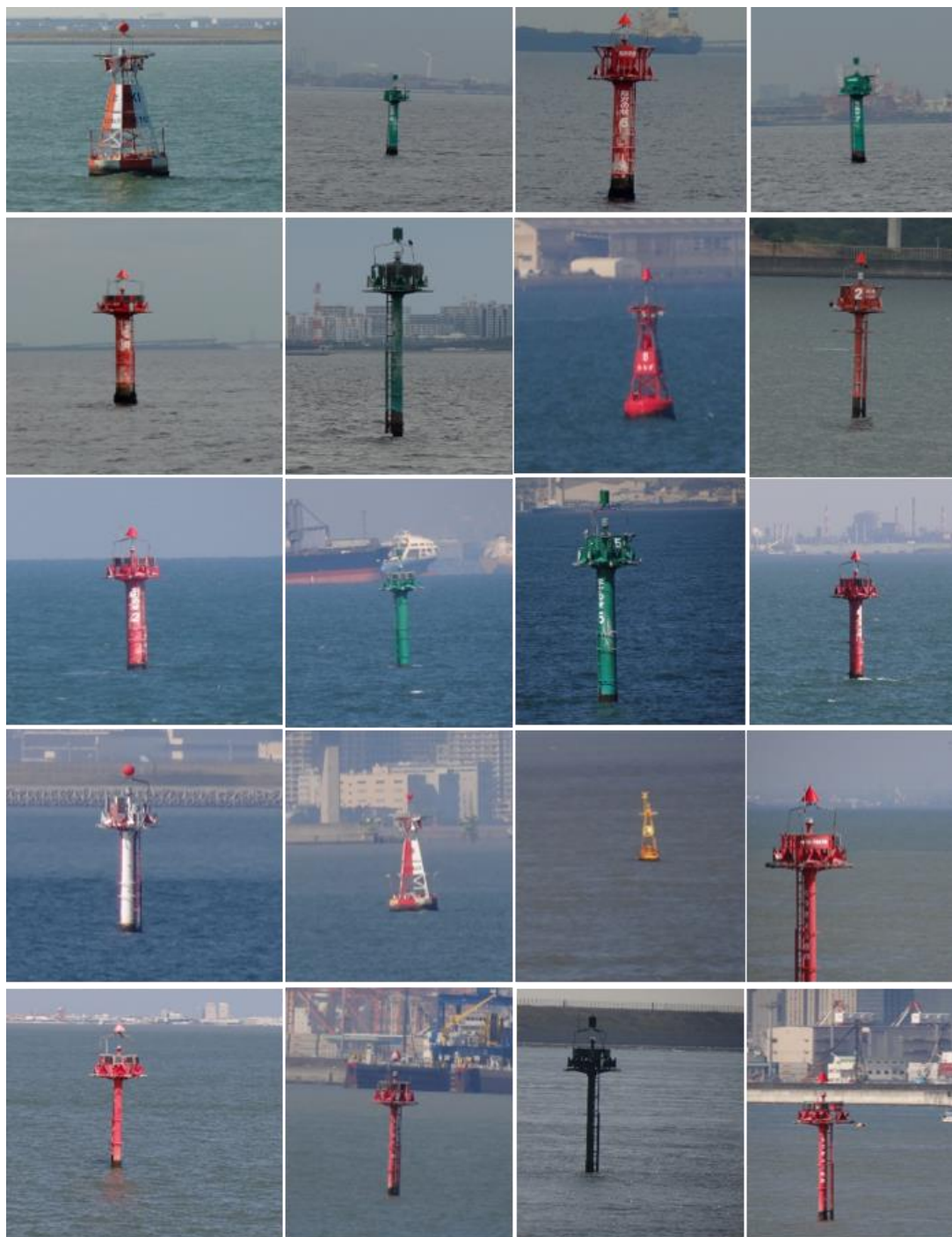


図 21.ブイの画像

7-2 データセット 1 を用いた静止画像の認識結果



図 22. 漁船の認識結果
100%確率の漁船 (s1)

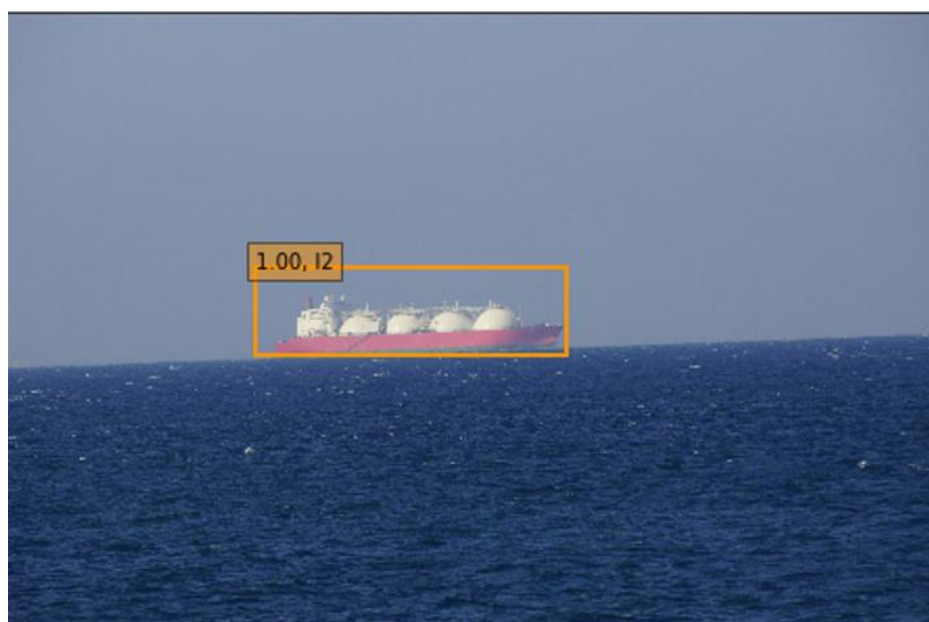


図 23. P C C の認識結果
100%確率の LNG 船 (l2)

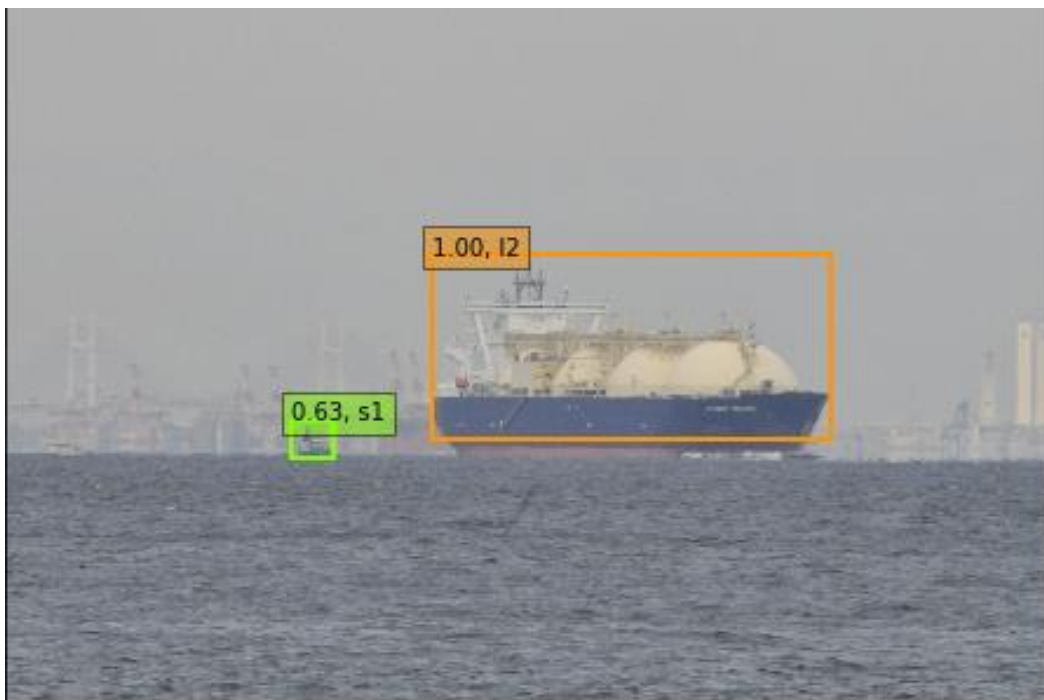


図 24. P C C と漁船の認識結果
63%確率の漁船 (s1) と 100%確率の LNG 船 (l2)



図 25. 漁船の認識結果
96%確率の漁船 (s1)



図 26. 漁船の認識結果
88%、64%、98%確率の漁船 (s1)

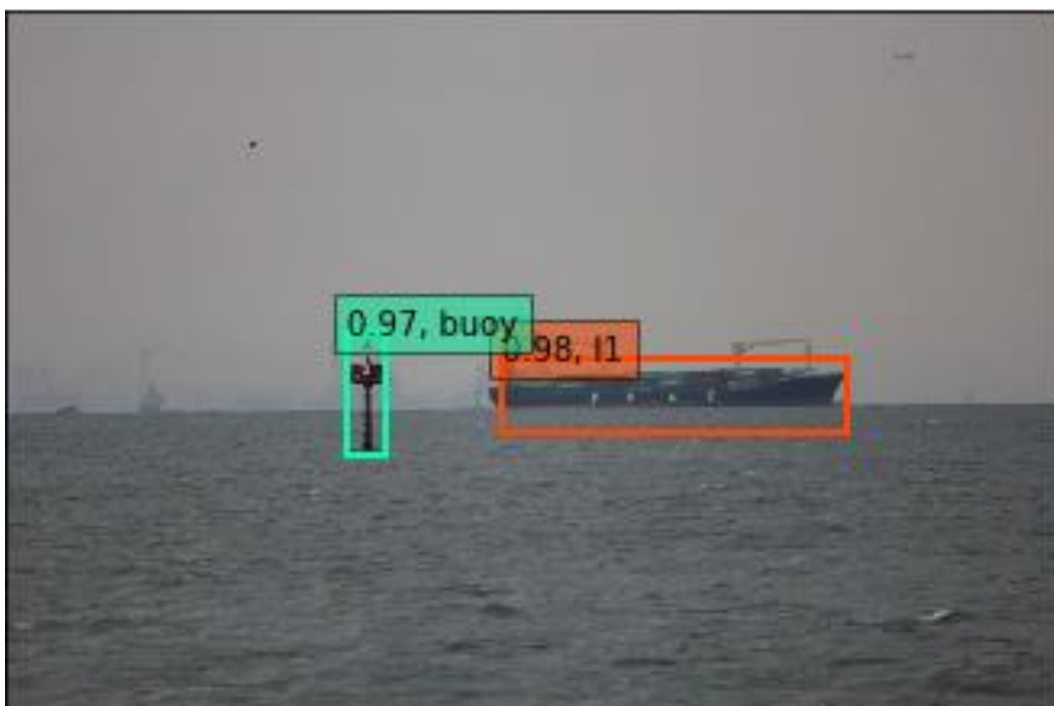


図 27. コンテナ船とブイの認識結果
97%確率のブイ (buoy) と 98%確率のコンテナ船 (11)

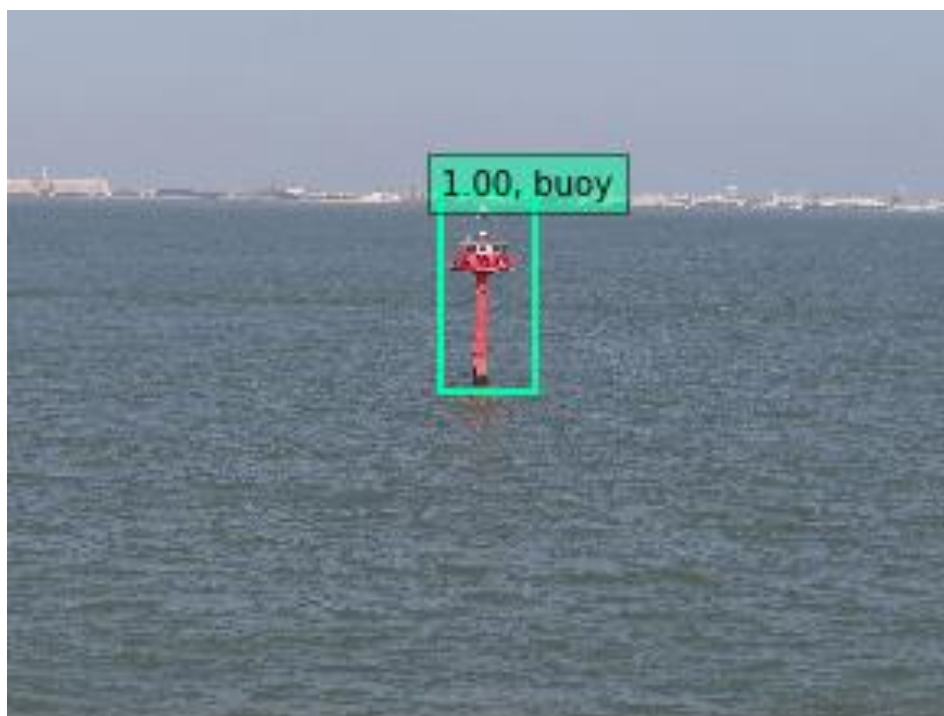


図 28. ブイの認識結果
100%確率のブイ (buoy)

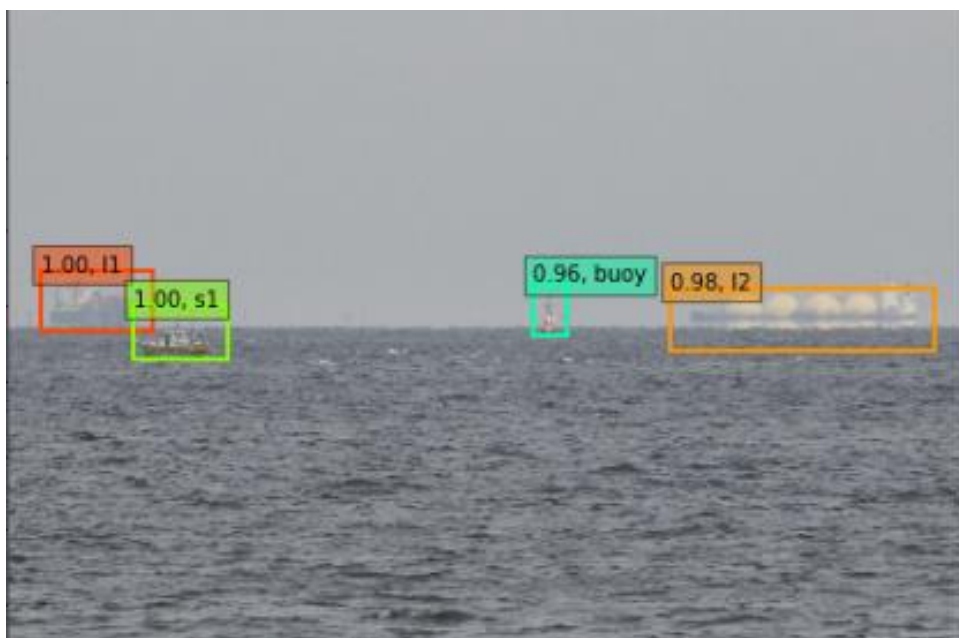


図 29. コンテナ船、PCC、漁船とブイの認識結果
100%のコンテナ船 (l1)、100%確率の漁船 (s1)、96%確率のブイ (buoy)、98%確率
の LNG 船(l2)

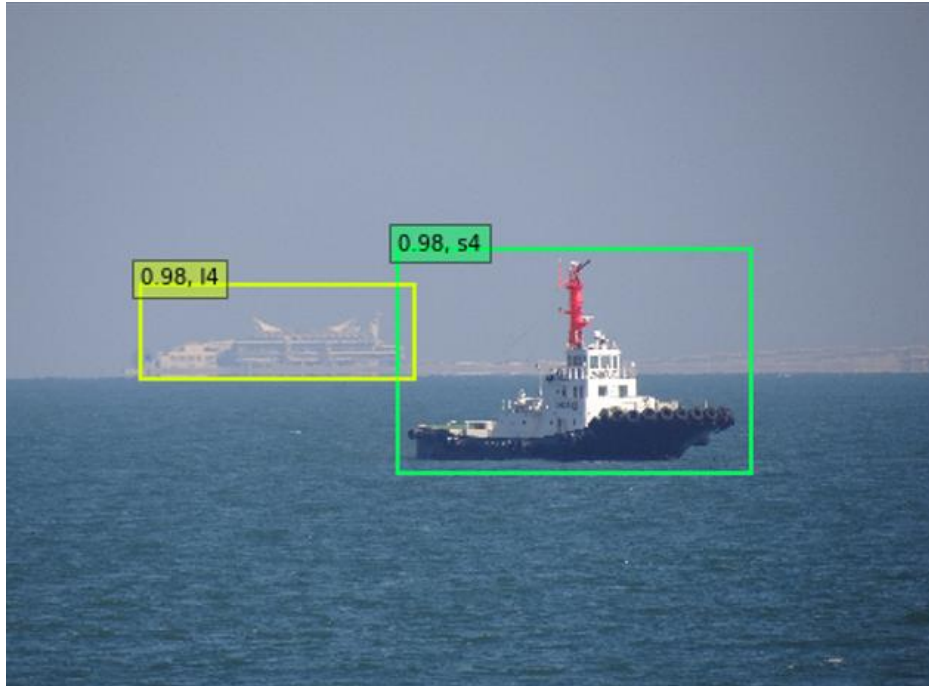


図 30. 他の大型船と他の小型船の認識結果
98%確率の他の大型船 (14) と 98%確率の他の小型船 (s4)
そのうちに、タグボートは他の小型船と認識された。これは誤分類である。



図 31. コンテナ船の認識結果
98%確率のコンテナ船 (11)

表 1 データセット 1 を用いた静止画像の認識率

認識率 (%) 距離(NM)	コンテナ船 (11)	LNG 船 (12)	PCC 船 (13)	他の大型船 (14)	漁船 (s1)	プレジャーボート (s2)	タグ (s3)	他の小型船 (s4)	ブイ (buoy)
0~5	100	99	88	99	100	100	99	99	100
5~10	99	98	86	91	92	86	85	94	89
10~	86	76	65	67	63	67	62	77	84

表 1 を見ると、

- 0~5 海里の認識率は、88~100%である。
- 5~10 海里の認識率は、86~99%である。
- 10 海里以上の認識率は、63~86%である。
- 大型船の認識率は小型船やブイなど小物体の認識率より高い。

何故なれば、学習画像は 1980*1980 ピクセルの画像である。SSD学習モデルは画像を読み込む前に、自動的に入力画像を 300*300 ピクセルの画像に変更する。そうすると、元画像中の小物体は消える可能性がある。学習するときに、小物体の学習率は下がると思う。

7-3 データセット 1 を用いた動画の認識結果

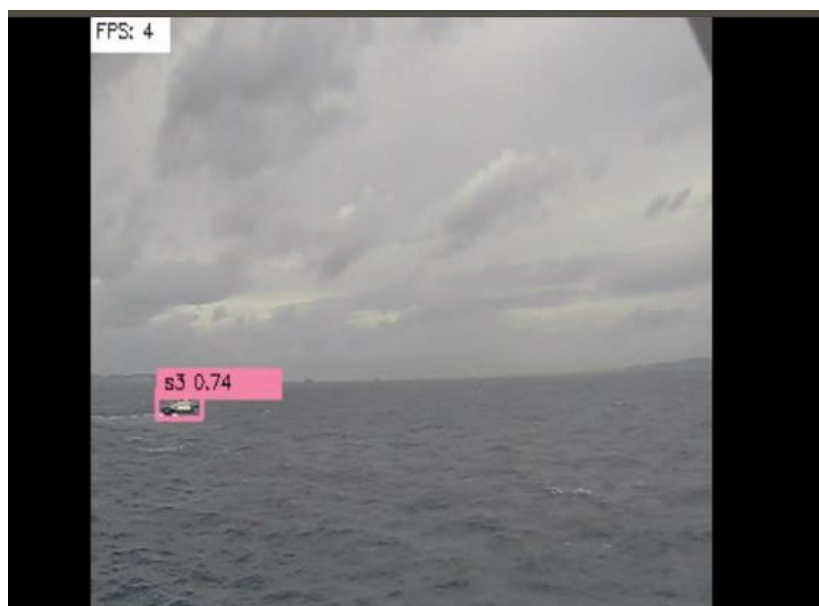


図 32. 動画に対して小型船の認識結果
74 % 確率のタグボート (s3)



図 33. 動画に対して大型船の認識結果
75%確率の漁船 (s1)
実際は他の大型船であるが、漁船と認識された。これは誤分類である。

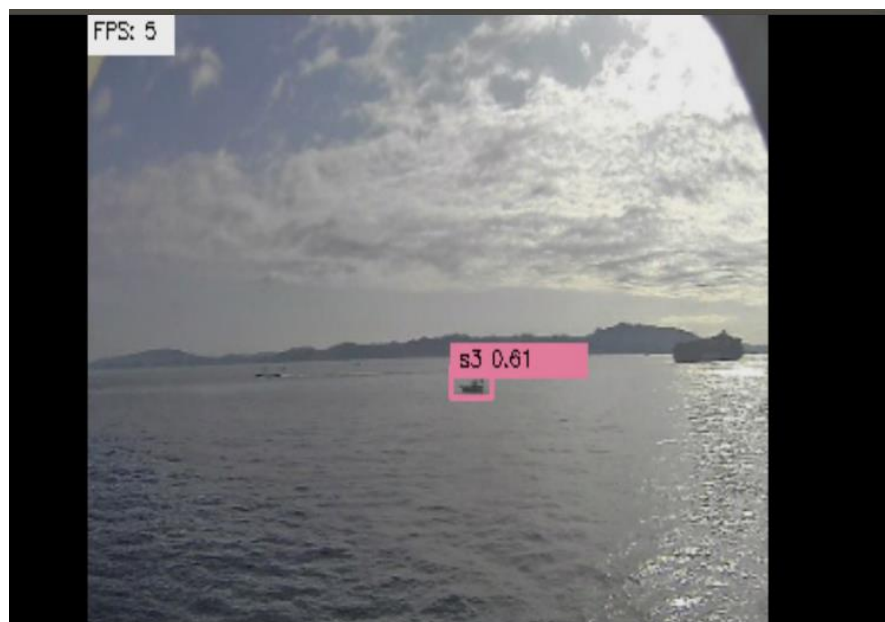


図 34. 動画に対して小型船の認識結果
61%確率のタグボート (s3)
実際は漁船であるが、タグボートと認識された。これは誤分類である。

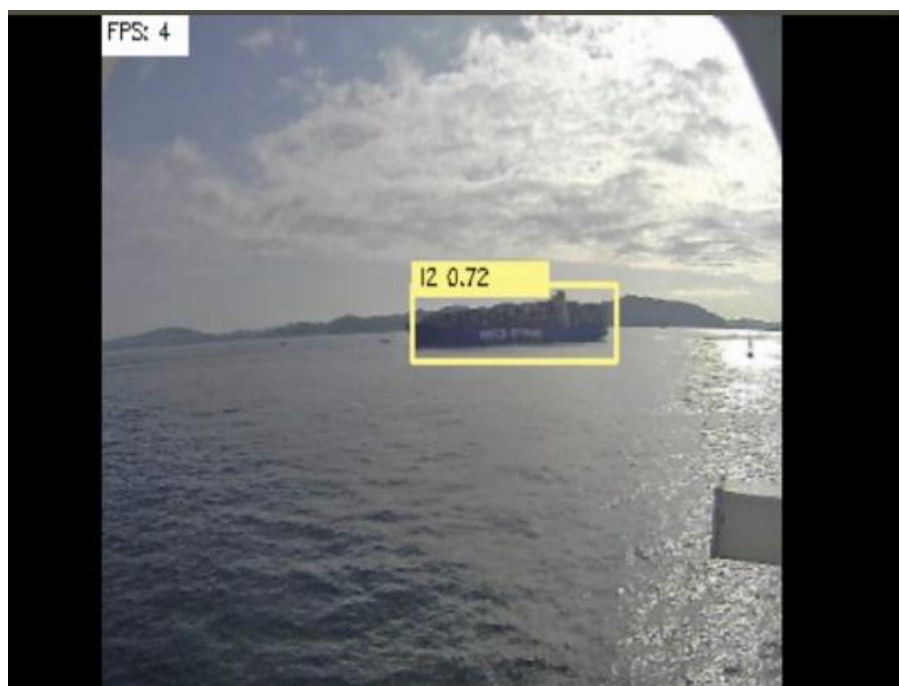


図 35. 動画に対して大型船の認識結果
72%確率の LNG 船 (12)
実際はコンテナ船であるが、LNG 船と認識された。これは誤分類である。

表2 データセット1を用いた動画の認識結果

認識率 (%) 距離(NM)	コンテナ 船 (11)	LNG 船 (12)	PCC 船 (13)	他の大 型船 (14)	漁船 (s1)	プレジャ ーボート (s2)	タグ (s3)	他の小 型船 (s4)	ブイ (buoy)
0~5	72	72	68	78	75	78	61	67	74
5~10	65	62	65	67	65	63	53	56	53
10~	50	43	34	54	45	43	32	43	56

表2を見ると、

- 0~5 海里の認識率は、61~78%である。
- 5~10 海里の認識率は、53~67%である。
- 10 海里以上の認識率は、34~56%である。
- 静止画像の認識率よりも動画の認識率が低い。

静止画像と動画を認識するために最優の 50epoch の weight をかけ、静止画像を用いた実験で 100%に近い結果を得た(表1)。ただ、動画の認識率はあまり良くなく、誤分類は多々存在している。航海中の船舶の位置はよく捉えられるが、太陽光や天候、波、島などの影響があり、分類が必ずしも正確ではない。同時に認識率を高めるために、データセットの分類精度を改善する必要がある。さらに、さまざまな状況や環境に対応するために、より多くの学習用画像を用意する必要がある。船の揺れは結果にも影響を与える。

上記の原因は航海環境の客観原因以外に、認識対象までの距離汐路丸と物標間の距離が離れているため、小型船やブイなどが小さく映ることも一つの原因になっている。また、SSDモデルが300*300ピクセルの画像を読み込むしかないと主原因となっている。SSDモデルの入力画像の大きさを変更できないから、学習用データセットの標注方式を変更しなければならない。認識精度の向上に向けて自船と他船の距離を考慮する必要がある。

7-4 データセット2

動画を用いた実験の認識精度の向上のために、データセットを汐路丸の IP カメラからもらった航海映像と取り替える。この航海動画を静止画像に変更し、約5,700枚の学習用データセット2を作った。汐路丸は他船の近くにいけないから、航海映像からの障害物がすごく小さい。この問題を解決するために、海平線の近くの障害物を5つのカテゴリ(11(近くの大型船)、12(遠くの大型船)、s1(近くの小型船)、s2(遠くの小型船)、buoy)に分類した。その後、データセットを約200epoch訓練し、最終的に200epoch

の weight を得た。プログラミングをこの weight でかけて障害物を認識する。

距離の問題を解決に向けて、水平線から下 100 ピクセル（汐路丸から 5NM 以内）を境に障害物を 5つのカテゴリに分類する。データセット 2 の作成基準は以下の図 36 に示す。

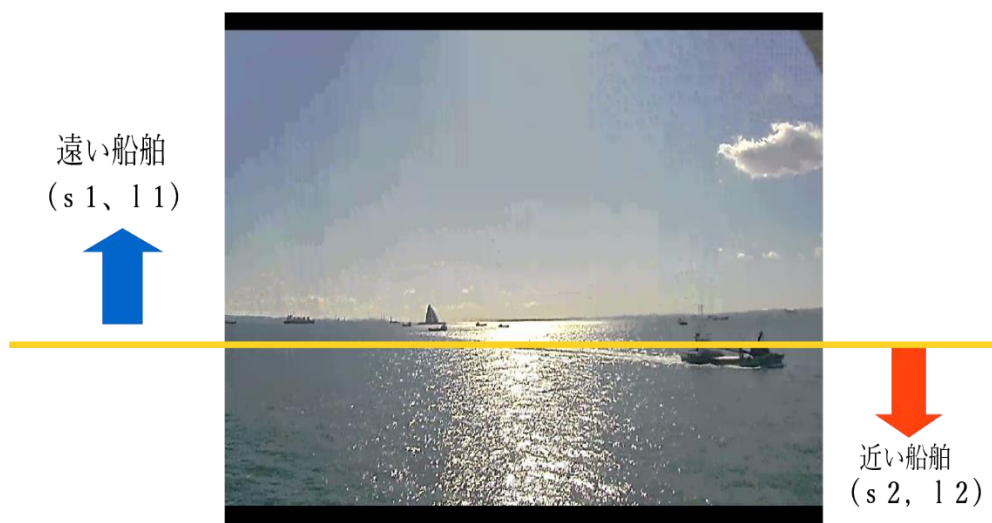


図 36. データセット 2 の作成基準

新しいデータセット 2 は以下に示す。

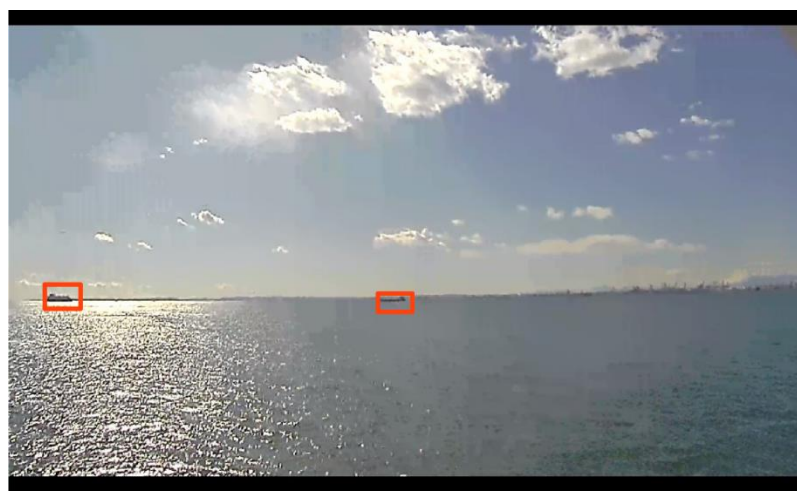


図 37. 遠くの大船舶の例

11 遠くの大船舶（漁船とプレジャーボート以外すべての船舶）



図 38 .近くの大型船舶の例

12 近くの大型船舶（漁船とプレジャーボート以外すべての船舶）



図 39. 遠くの小型船舶の例

s1 遠くの小型船舶（漁船とプレジャーボートのみ）



図 40. 近くの小型船舶の例
s2 近くの小型船舶（漁船とプレジャーボートのみ）



図 41. ブイの例
buoy（距離は関係なくすべての buoy）

7-5 データセット 2 を用いた静止画像の認識結果



図 42. 静止画像の認識結果

100%確率の遠くの大型船 (l1)、61%確率のブイ (buoy)
89% 確率の遠くの大型船 (l1)、100%確率の遠くの小型船 (s1)



図 43. 静止画像の認識結果

78%確率のブイ (buoy)

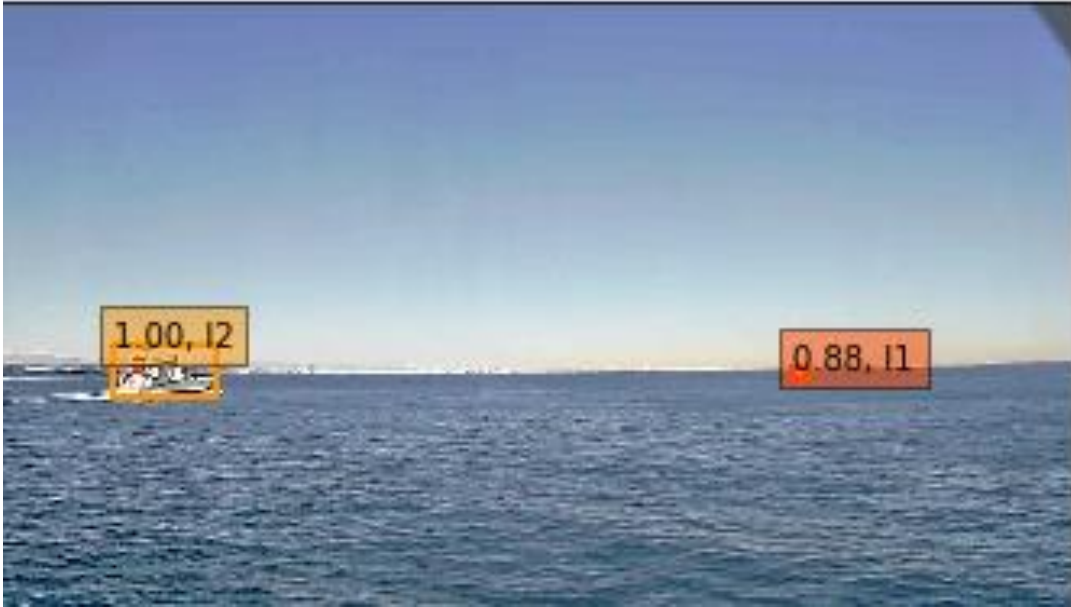


図 44. 静止画像の認識結果
100%確率の近くの大型船 (12) と 88%確率の遠くの大型船 (11)



図 45. 静止画像の認識結果
100%確率の近くの大型船 (12)



図 46. 静止画像の認識結果
95%確率のブイ (buoy)

表 3. データセット 2 を用いた静止画像の認識率

データセット 2	遠くの大型 船 (11)	近くの大型 船 (12)	遠くの小型 船 (s1)	近くの小型 船 (s2)	ブイ (buoy)
認識率 (%)	92	97	91	89	94

表 3 を見ると、

- データセット 2 を用いた静止画像の認識率は、89～97%である。
- 大型船の認識率、小型船やブイなどの小物体の認識率も向上した。

以上の図により、データセットを取り替えてから、静止画像を用いた実験の認識率を明らかに上がった。船舶やブイなどの障害物が小さくても認識できる。

7-6 データセット 2 を用いた動画の認識結果

乗船する前に、研究室に汐路丸から撮影された映像を利用して船舶認識の簡易実験を行った。データセット 2 を用いた動画の実験結果は以下に示す。



図 47. 動画の認識結果

62%確率の近くの大型船 (12)

実際は遠くの小型船であるが、近くの大型船と認識された。これは誤分類である。



図 48. 動画の認識結果

74%確率の近くの小型船 (s2)

実際は遠くの小型船であるが、近くの小型船と認識された。これは誤分類である。



図 49. 動画の認識結果

74%、82%確率の近くの大型船 (l2)

実際は遠くの大型船であるが、近くの大型船と認識された。また、未検出がある。

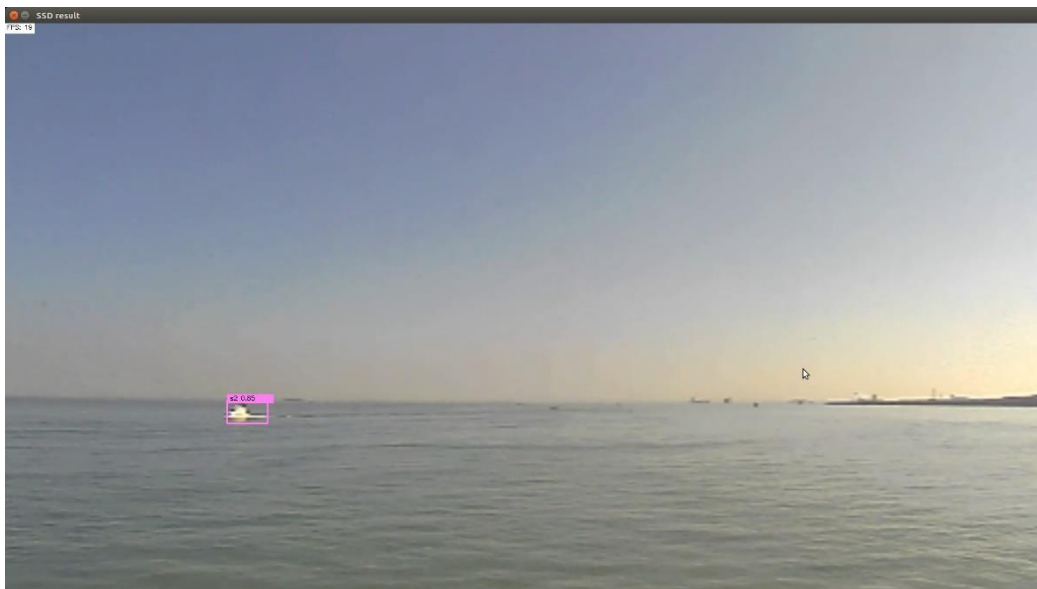


図 50. 動画の認識結果

85%確率の近くの小型船 (s2)

これは正確な認識結果である。

表 4. データセット 2 を用いた動画の認識率

データセット 2	遠くの大型 船 (11)	近くの大型 船 (12)	遠くの小型 船 (s1)	近くの小型 船 (s2)	ブイ (buoy)
認識率 (%)	64	73	43	61	78

表 4 を見ると、

- 動画像では認識率に大きな変化（改善）がない。
- 誤検出と未検出が多々存在している。
- 船舶やブイなどの障害物の位置が正確に表示されているが、カテゴリの分類が正しくない。

静止画像に対する認識率は 90% に近い結果を得たが、動画像に対する認識率は向上していない。動画像の認識実験では、同一の物体を連続して検出できない。異なるラベル（物体）として認識することが多々ある。

8、実験

8-1 実験環境

2018年1月16日から18日までの3日間、汐路丸に乗船してSSDの船舶認識実験を行った。また、学習データセットを変えて、2月6日から8日までの3日間もSSDの船舶認識実験を行った。

実験環境は以下の表4に示す。

表5. 実験環境

実験日	2018年1月16日～18日 2018年2月6日～8日
実験海域	東京湾 館山湾
実験用船	汐路丸
実験時間	日出後～日没前
認識対象	汐路丸の周囲の船舶
実験機材	汐路丸に設置されたIPカメラ コンピュータ 一台

実験用コンピュータの環境配置

学習モデル：SSD (keras-Tensorflow)

OS：Ubuntu OS 64bit

GTX1070

CUDA 8.0

cuda 5.1

Python 2.7

Anaconda 4.3

Tensorflow-gpu 1.4

Keras 2.1.2

Opencv 3.4.0

本実験はSSDの学習モデルを使って、汐路丸の周囲の船舶を認識することを目標とした。また、使った重みは70epoch学習された重みである。

8-2 航行中の船舶から取得した映像による実験

2018年1月16日～18日と2018年2月6日～8日この二次の実験は東京湾と館山湾の

船舶の認識実験を行った。全部の学習データセットは 7000 枚ぐらいがあり、s1、s2、s3、s4、11、12、13、14、buoy 10 カテゴリを分類される。実験結果はそれぞれ以下に示す。
2018 年 1 月 16 日～18 日の実験結果



図 51. 2018 年 1 月 16 日～18 日航海実験の実験結果



図 52. 2018 年 1 月 16 日～18 日航海実験の実験結果



図 53. 2018 年 1 月 16 日～18 日航海実験の実験結果



図 54. 2018 年 1 月 16 日～18 日航海実験の実験結果

2018年2月6日～8日の実験結果



図 55. 2018年2月6日～8日航海実験の実験結果

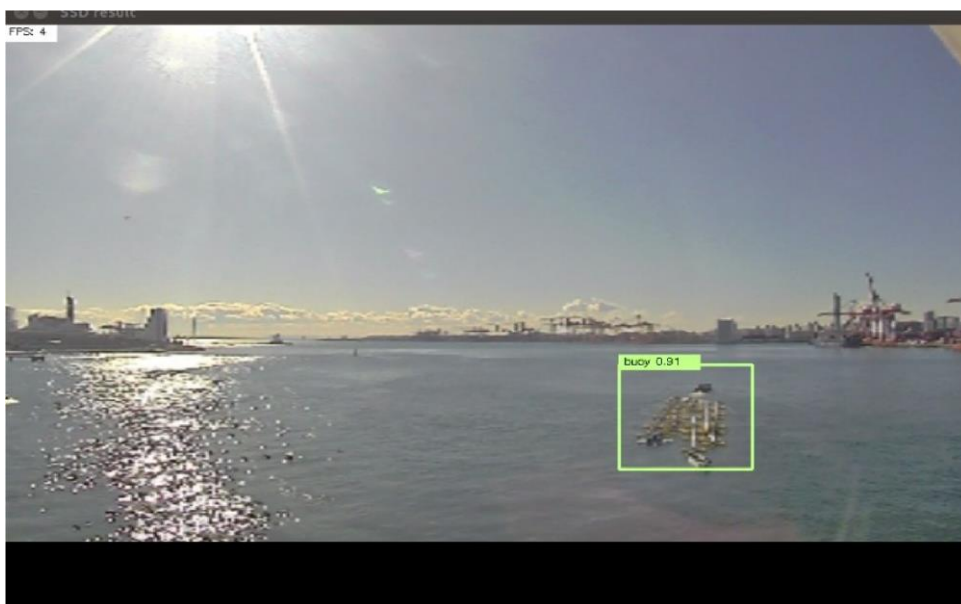


図 56. 2018年2月6日～8日航海実験の実験結果



図 57. 2018 年 2 月 6 日～8 日航海実験の実験結果

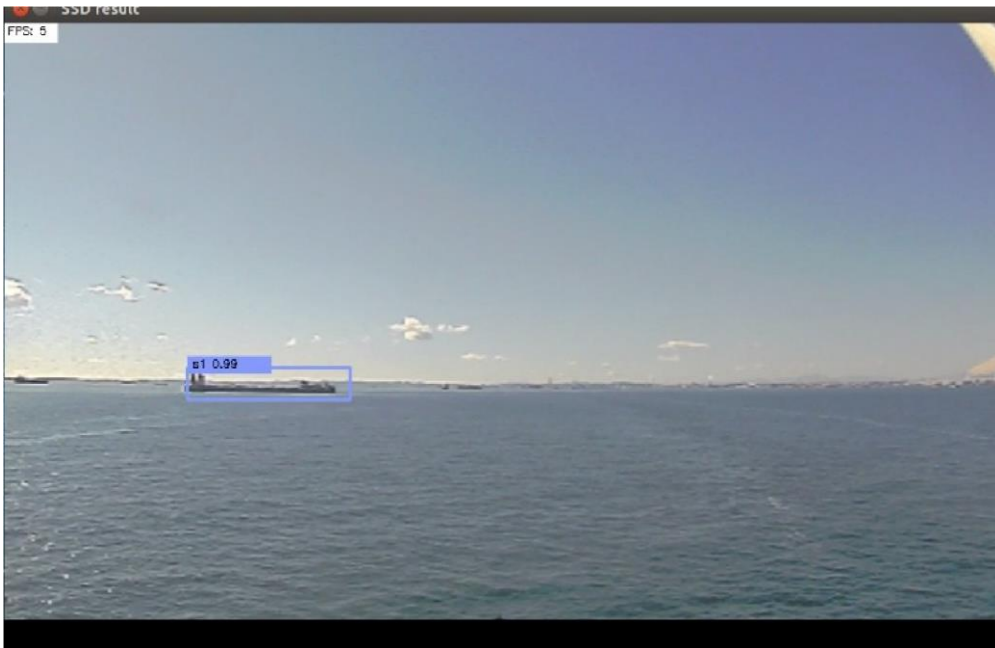


図 58. 2018 年 2 月 6 日～8 日航海実験の実験結果



図 59. 2018 年 2 月 6 日～8 日航海実験の実験結果



図 60. 2018 年 2 月 6 日～8 日航海実験の実験結果

以上の結果を示すように、他船やブイなどの障害物の位置情報は正しく表示されているが、カテゴリの分類は正しくない。また、静止画像を用いた実験は100%に近い結果を得たが、映像を用いた実験はよい結果を獲得していない。そして、映像に対する船舶認識の精度の向上は必要である。データセットを変更した後、静止画像に対して認識率が90%に近い結果を得た。静止画像に対して認識率が上がったが、航海環境の客観原因が存在しているので、静止画像に対して認識はまた10%の未検出と誤検出の場合がある。動画に対して認識率は明らかな変化がない。誤検出と未検出の場合が結構ある。船舶やブイなどの障害物の位置が正確に表示されているが、カテゴリの分類はほとんど正しくない。そして、SSDモデルが300*300ピクセルの画像を読み込むしかないことは主な原因となっている。

この問題を解決するために、時間関係の時系列を用いて、認識率の向上と未検出の削減と考える。

9、時間情報を利用した認識精度の向上

9-1RNN

RNN (recurrent neural network) [35]は再帰型ニューラルネットワークとも呼ばれ、自然言語処理の分野で高い成果がある。また、RNNは回帰結合入力の深層ニューラルネットワークである。伝統的なFNN(フィードフォワードニューラルネットワーク)とは異なり、RNNは入力間の文脈的相関を扱うことができる方向ループを導入する。RNNは文章など連続的な情報を利用でき、数値の時系列データなどの系列データのパターン認識するように設計されたニューラルネットワークのモデルである。伝統的なニューラルネットワーク(CNN)は時系列なデータを記憶できないが、RNNは時系列なデータを用いて、次の出力を予測できる。時系列学習モデルではそれぞれの入力の時系列データで、それぞれの処理の間も関連性を持っている。RNNの目的は、系列データを処理することである。従来のニューラルネットワークモデルでは、入力層から隠れ層、出力層に各層が完全に接続され、各層間のノードは接続されていない。しかし、RNNでは、系列の現在の出力も以前の出力に関連している。具体的な表現形式は、ネットワークが以前の情報を記憶し、それを現在の出力の計算に適用する。すなわち、隠れ層の間のノードは接続されている。また、隠れ層の入力は入力層の出力だけでなく、前の時点の隠れ層の出力も含まれる。理論的には、RNNは任意の長さの系列データを処理することができる。RNNでは入力されるデータは時系列データと呼ばれる。たとえば、時系列データ： x_1, x_2, x_3, x_4 があり、これらの時系列データを実際の分野で活用する。自然言語処理のモデルなら、 x_1 は最初の単語、 x_2 は2番目の単語などと見なすことができる[34]。音声処理のモデルなら、 x_1, x_2, x_3, \dots は各フレームの音声信号である[35]。

9-1-1RNN ネットワークの構造図

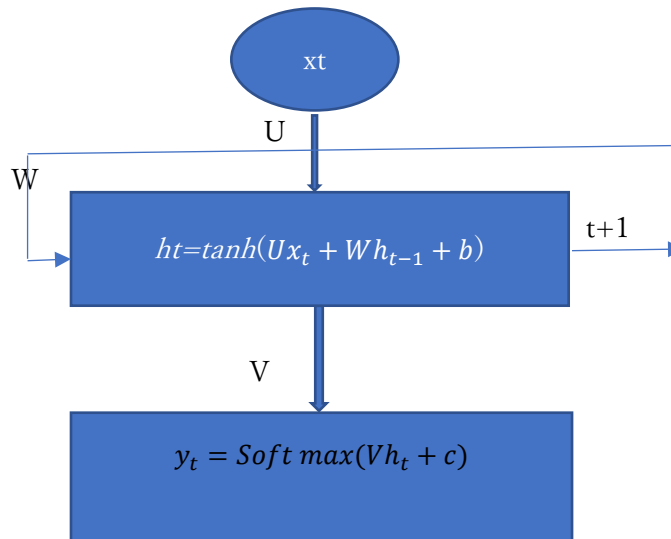


図 61. RNN ネットワークの構造図

伝統な RNN 構造は多個入力と多個出力がある。時系列データは、元のニューラルネットワークのように処理できない。時系列問題をモデル化するために、RNN は隠れ状態 h の概念を導入し、時系列のデータから特徴を抽出し、それを出力に変換することができる。

- x_t は入力層で時系列データの入力を処理する。
- h_t は隠れ層の処理である。動画の各フレームの分類ラベルと位置情報を計算する。計算は各フレームに対して実行されるので、入力シーケンスと出力シーケンスは同じ長さになる。
- y_t は出力層である。 h_t と h_{t-1} からもらった画像の特徴を softmax 処理で実行する。また、それぞれの処理の間は固定な重なりついている。CNN の場合ではそれぞれの処理の間の重なりが固定ではない。

9-1-2 多層 RNN 構造 (N vs 1)

もし出力は序列データではなく、単独な数値である場合は

$$h_t = f(Ux_t + Wh_{t-1} + b) \quad (9-1-1)$$

$$y = \text{Soft max}(Vh_t + c) \quad (9-1-2)$$

9-1-3 多層 RNN 構造 (1 vs N)

$$h_t = f(Ux + Wh_{t-1} + b) \quad (9-1-3)$$

$$y_t = \text{Soft max}(Vh_t + c) \quad (9-1-4)$$

ここで、 f は、一般に \tanh または ReLU のような非線形活性化関数である

9-1-4 Attention モデル

隠れた状態 h (hidden state) をモデルメモリと考えることはできるが、長い情報を保存することはできない。

従来のニューラルネットワークモデル (各層では異なるパラメータを使用) と異なり、RNN は各繰り返しで同じパラメータセット (U 、 V 、 WU 、 V 、 W) を共有する。たとえば、文の感情を予測する場合、RNN の最終出力を使用してすべての単語を入力するのではなく、予測を行うことができる。このモデルの主な特徴は、時系列のすべての情報が RNN の非表示状態に格納されていることである。隠れ層状態 h_t をネットワークの記憶セルと見ることができる。 h_t には、前のすべてのステップの隠れ層状態が含まれている。実際には、ネットワークの複雑さを減らすために、 h_t にはすべてのステップではなく、前のあるステップの隠れ層状態のみが含まれている。

9-1-5 RNN に基づいて物体認識の研究

2016 年に Bell S ら [36] は skip pooling と RNN layer を利用し、IRNN というニューラルネットワークを提案した。修正された IRNN は、LSTM に似た高速性と精度を持っている。VOC2012 では、主にマルチスケールの feature map と layer の使用により、認識精度が 76.4% に達した。RoI pooling は、マルチスケールの feature map 上で行われるため、各スケール上の feature map のレスポンスの大きさは同じではないため、L2-norm の関数を行ってから RoI feature によってスケールを連結する必要がある。次に、スケール (ネットワークによって学習されたスケール) を統一する。最後の feature map は、RNN によって取得される。ボトルやプラントなどの小物体の認識効果がよくない。初期の feature map で識別するほうが良いと思う。最も早い feature scale 情報がよりよく保存される。RNN は特徴の抽出のみに使っている。物体の時間序列を考慮していない。一回目で物体を認識したら、Attention モデルを用いて、これらの object を feature map から削除し、小物体を識別するのみほうが良い。

また、Jiang Wang ら [37] は画像の複数ラベル識別ができる学習モデル CNN-RNN を提案した。この論文の主な目的は、画像の複数ラベル識別を行うことである。文章から when using the same image features to predict multiple labels, objects that are small in the images are easily get ignored or hard to recognize independently. そういう一言がある。同一の feature map を用いて複数ラベルを予測する際に、小さな物体を無視する場

合がある。したがって、2つの並列ネットワークを利用し、二つ目のネットワークに入力するときに、最初にラベルの埋込み (label embedding) を取得し、RNN からのベクトルと feature map からの画像の埋込み (image embedding) を融合することによって画像の次のラベルを出力する。この論文に対する思考は Deconvolution を使用し、画像と同じサイズに feature を拡張することで、小さな物体を識別できると考えている。また、異なるサイズの畳み込みカーネルを使用するほうが良いと考えている。

9-2LSTM

LSTM (Long Short Term) ネットワーク構造

簡単な RNN で現実的にはかなり前の古い情報を考慮用に学習できない。もし学習情報の間隔が長くなると、勾配消失の問題がある。LSTM[38]は従来の RNN セルでは長期依存が必要なタスクを学習することができなかった問題を解決したモデルである。LSTM は入力ゲート、出力ゲートと忘却ゲートという情報の取捨選択機能がある。

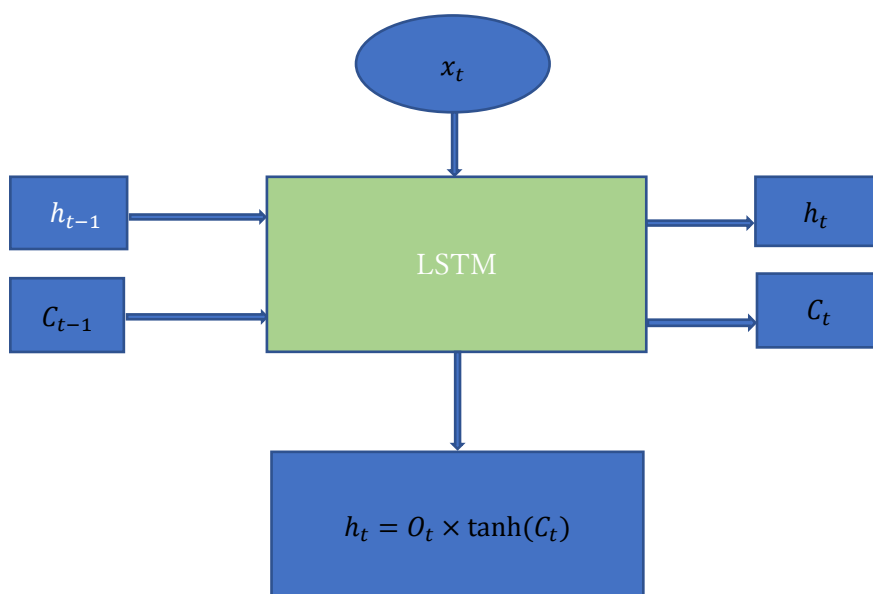


図 62. LSTM ネットワークの構造

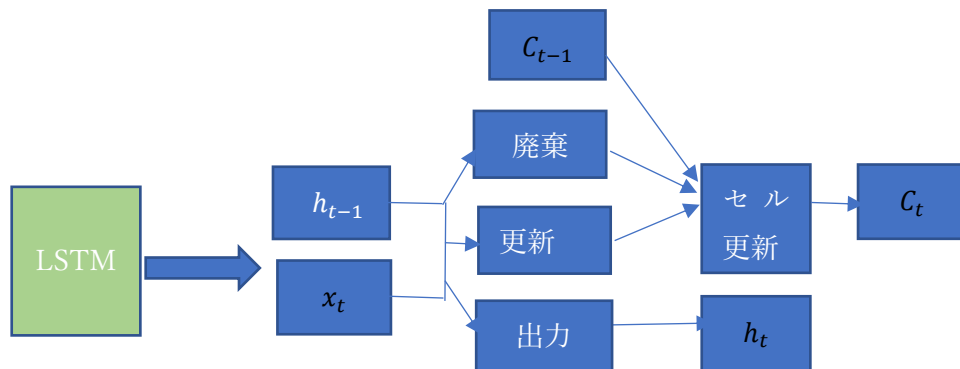


図 63. 取捨選択機能

9-2-1 忘却ゲート

LSTM は構造の上で RNN と大体同じであるが、四つのニューラルネットワークがあり、それぞれは廃棄、更新、出力、セル更新部分である。

- 廃棄

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (9-2-1)$$

- 更新

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (9-2-2)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (9-2-3)$$

- セルの状態を更新する

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (9-2-4)$$

- 情報を出力する

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (9-2-5)$$

$$h_t = o_t * \tanh(C_t) \quad (9-2-6)$$

9-2-2 今流行している LSTM モデル

多様な LSTM 変種の 1 つは、2000 年に Gers と Schmidhuber[39] によって提案されたもので、「peephole connection」が追加されている。つまり、ドアレイヤーがセル状態の入力を受け入れるようにする。

$$f_t = \sigma(W_f[C_{t-1}, h_{t-1}, x_t] + b_f) \quad (9-2-7)$$

$$i_t = \sigma(W_i[C_{t-1}, h_{t-1}, x_t] + b_i) \quad (9-2-8)$$

$$o_t = \sigma(W_o[C_t, h_{t-1}, x_t] + b_o) \quad (9-2-9)$$

Peephole connection

別の LSTM 変種は、coupled を使って忘れてゲートに入ることである。以前に忘れられたものと新しい情報を追加する必要があることを分ける決断とは異なり、ここでは一緒に決定する。

$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t \quad (9-2-10)$$

Coupled 忘却ゲートと入力ゲート

9-2-3 RNN と LSTM の比較

入力ゲート、忘却ゲートと出力ゲートの概念を導入したため、RNN に関する勾配消失の問題を解決できる。LSTM は従来の RNN セルでは長期依存が必要なタスクを学習することができなかつた問題を解決できる。

9-2-4 LSTM に基づいて物体認識の研究

Xiaodan Liang ら[40]から Graph LSTM 学習モデルを提案している。この論文の主な考え方は、元々のポイント・バイ・ポイント・シーケンスを捨て、スーパーピクセルを序列として使用することである。スーパーピクセルの入力順序は、前の FCN によって得られた信頼マップ(confidence map)に基づいて決定される。Graph LSTM は、セマンティックセグメンテーションに応用されている。物体認識に応用できるかどうかかわからない。

9-3GRU

GRU[41]は、LSTM をもう少しシンプルにしたモデルで、LSTM の変種である。入力ゲートと忘却ゲートを「更新ゲート」として1つのゲートに統合している。

LSTM と同様に、このような忘却・更新ゲートを導入することで、長いステップ前の出来事の特徴の記憶を維持しやすくなる。そして、学習中に誤差を容易に逆伝播することができ、勾配消失の問題を軽減することになる。また、セル状態と隠れ状態、およびその他の変更が混在している。最終的なモデルは標準的な LSTM モデルよりも簡単で、非常に一般的な変形である。

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (9-3-1)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (9-3-2)$$

$$\tilde{h}_t = \tanh(W[h_t * h_{t-1}, x_t]) \quad (9-3-3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (9-3-4)$$

GRU(Gated Recurrent Unit)ネットワーク構造

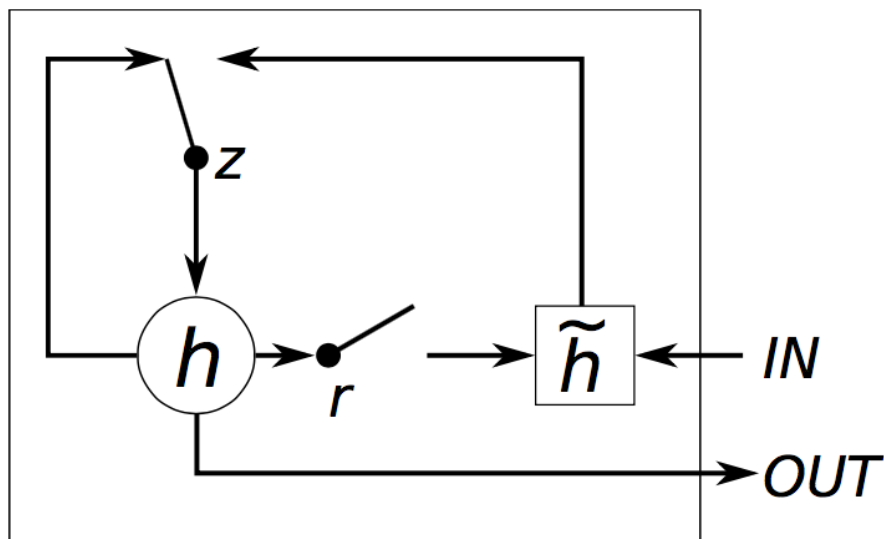


図 64. GRU ネットワークの構造[41]

- z_t は更新ゲートであり、活性化 h_t を更新するときの論理ゲートである。
- r_t はリセットゲートであり、候補活性化を決定するとき、以前の活性化 h_t を放棄するかどうかを判断するゲートである。
- \tilde{h}_t は、候補活性化であり、 $[x_t, h_{t-1}]$ を受け取る。
- h_t は活性化であり、これは GRU の隠れ層であり、 $[h_{t-1}, \tilde{h}_t]$ を受け取る。

9-3-1 LSTM と GRU の比較

LSTM と GRU との間にどれがいいかと言えない。GRU のパラメータはより少ないので、学習速度が速く、必要なサンプル数が少なくなる。LSTM はより多くのパラメータを持ち、多数のサンプルを持つ状況に適しており、より良いモデルを得ることができる。

9-4QRNN

QRNN(Quasi-Recurrent Neural Network) [42] は従来の RNN の学習効率の向上に向けて提案された。一般的な RNN では、一方向に過去の状態から遷移しないと、状態が計算できないために、並列化しにくく学習に時間がかかる。QRNN では、CNN の仕組みを活かし、並列性を確保することで、高速かつ高精度の特性がある。

RNN は時間依存性があり、並列度が制限される。CNN は制限された受容野の影響を受けるため、情報伝達が遅すぎる。QRNN は、RNN と CNN の利点を統合し、自分の欠陥を回避しようとしている。

9-4-1 アルゴリズム

LSTM の計算式に続いて、QRNN のゲートの式は次のとおりである。

$$z_t = \tanh(W_z * x_t) \quad (9-4-1)$$

$$f_t = \sigma(W_f * x_t) \quad (9-4-2)$$

$$o_t = \sigma(W_o * x_t) \quad (9-4-3)$$

*は一次元畳み込み演算を表す。このステップ動作は時間序列に依存しない。

時間序列関係をモデル化するために、メモリは次のように異なる時間に更新される。

$$C_t = f_t * C_{t-1} + (1 - f_t) * z_t \quad (9-4-4)$$

$$h_t = o_t * C_t \quad (9-4-5)$$

このステップは時間に依存するが、計算量は少なく、要素単位の操作だけが使用される。

また、dropout を使用したソリューションも提案している[42]。

$$f_t = 1 - \text{dropout}(1 - \sigma(W_f * x_t)) \quad (9-4-6)$$

9-4-2 RNN,LSTM,GRU と QRNN の比較

QRNN は、演算行列の乗算演算に時間依存性がないため、適切に実装された QRNN は、LSTM / GRU よりも速度上の優位性を持っている。QRNN 本体は畳み込み演算であるため、ループよりも学習が容易である。

9-5LSTM、GRU に関する実験

映像を用いた検出精度の向上のために、LSTM を SSD に追加することを検討した。LSTM は時系列データを用いた学習モデルであるから、普通のデータセットを時系列データへ変換しなければならない。しかし、画像のデータセットを時系列データへ変換することが難しい。そして、本研究では、LSTM を SSD のネットワーク層の最後、予測の前に追加する。SSD ネットワーク層からの特徴を LSTM に渡し、予測の前にもう一度特徴を抽出、選択する。最後に学習済みモデルの重みを得る。その重みを利用して、映像の物体認識に応用する。

SSD-LSTM 学習モデルのネットワーク構造

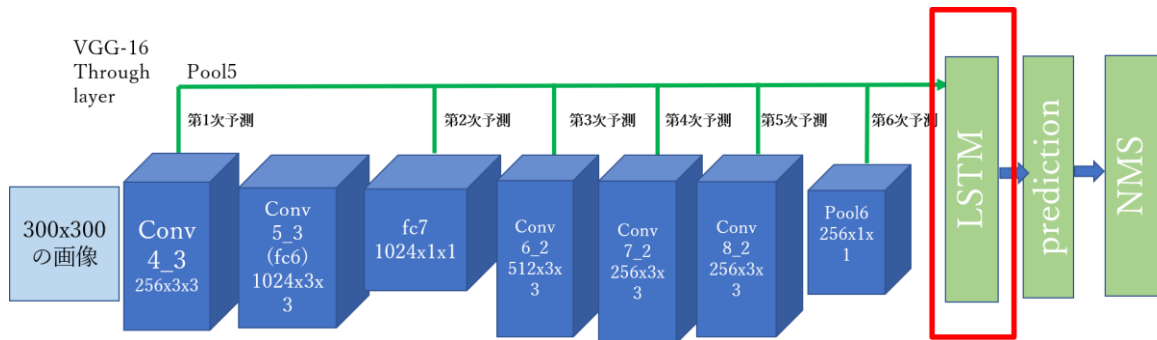


図 65. SSD-LSTM 学習モデルのネットワーク構造

9-6 SSD-LSTM に関する結果

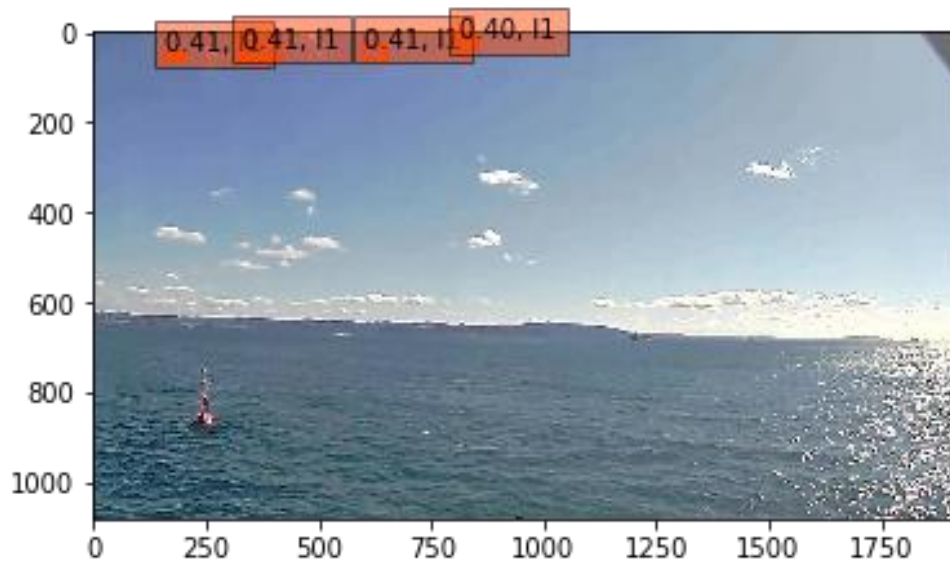


図 66.ブイの認識結果

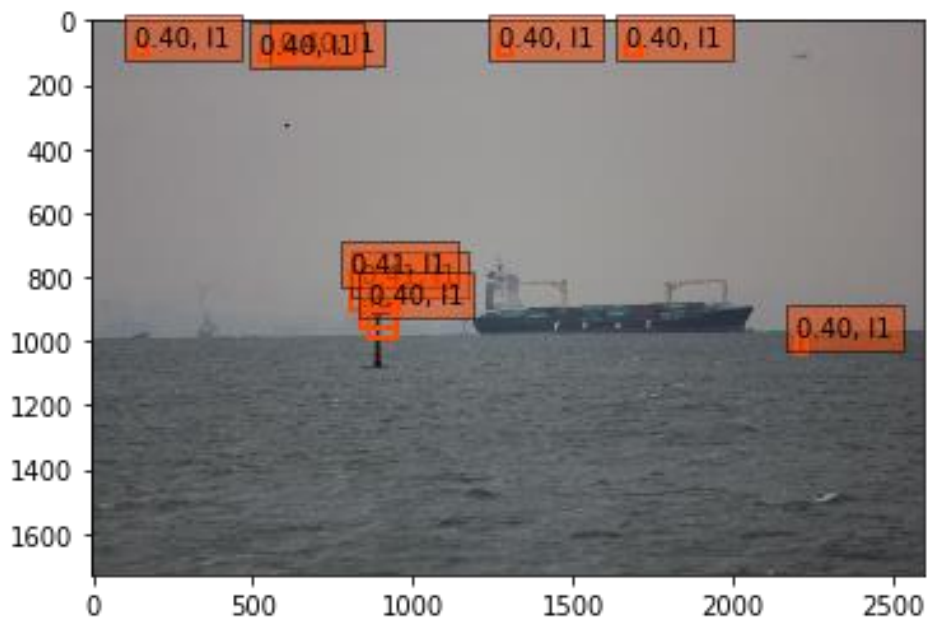


図 67. ブイと遠くの大型船舶の認識結果

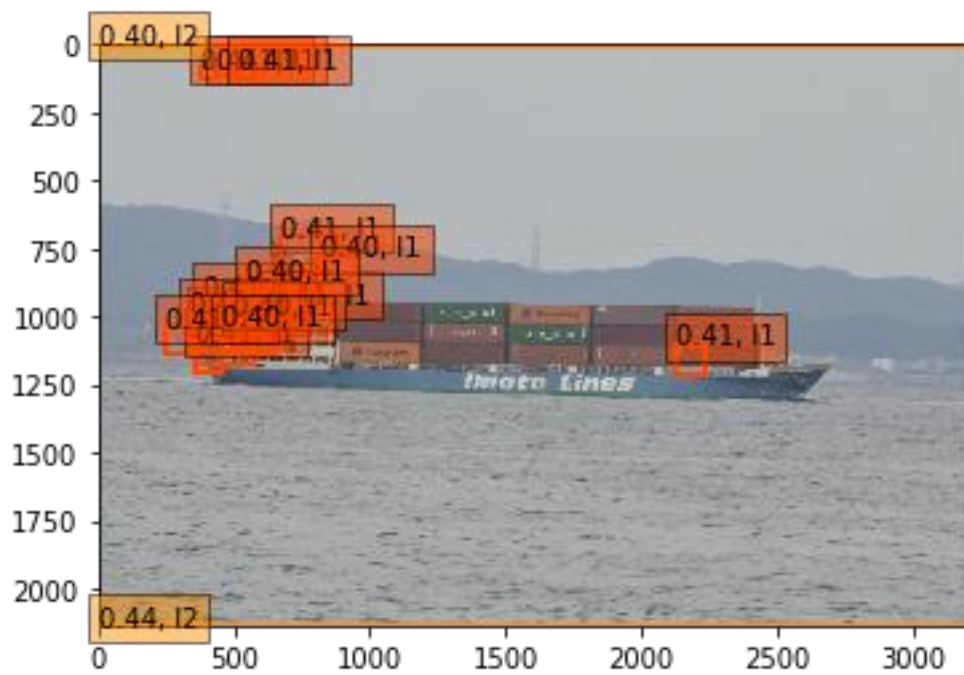


図 68. 近くの大型船舶の認識結果

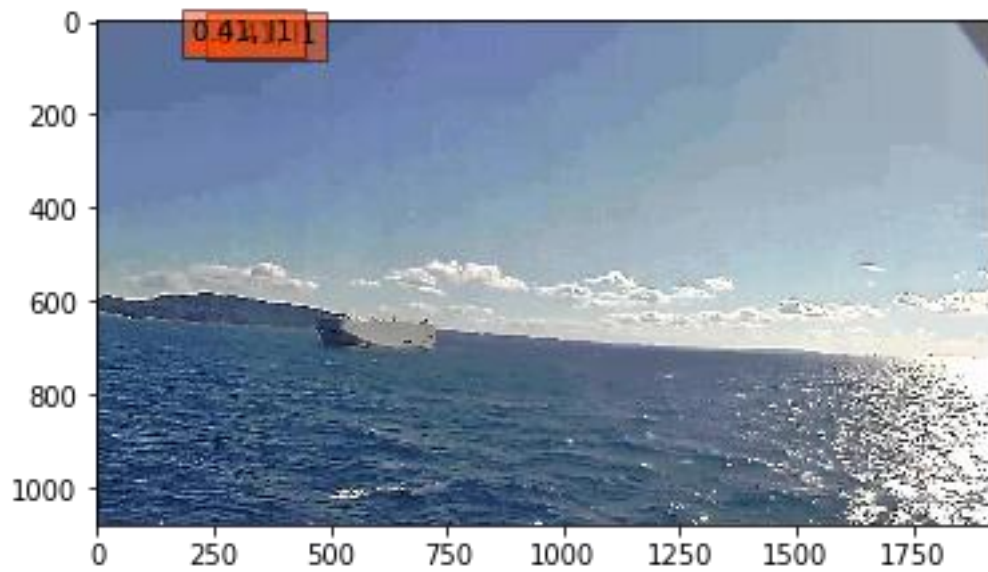


図 69.近くの大型船舶の認識結果

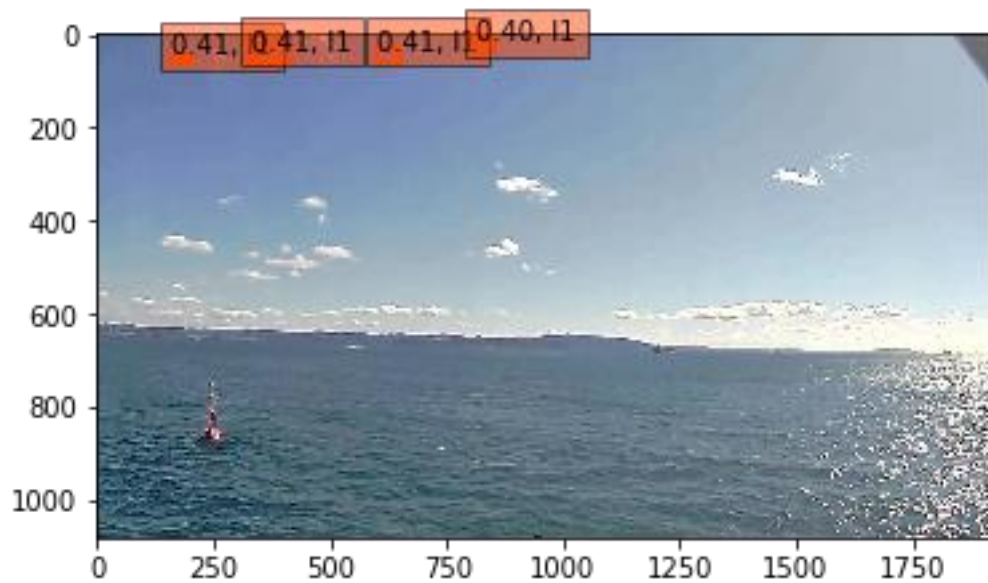


図 70.ブイの認識結果

SSD のネットワークに LSTM を追加した後、静止画像の認識結果は大幅に下がった。枠が出てきたが、正しくない。映像の船舶認識も行ったが、何にも検出していない。また、映像の表示はすごく遅く、毎秒 1 FPS になっている。

10、評価

海難事故では主な原因が見張り不十分なので、衝突がよく発生している船舶の認識実験を行った。海上航行映像からコンテナ船、LNG 船、PCC、その他の大型船、漁船、プレジャーボート、タグボート、その他の小型船、ブイの9のカテゴリのデータセットを作成した。このデータセットを約 50epoch を学習し、船舶認識を行った。静止画像の方は良い結果を取得した。大型船の認識率は 100%に近い結果を得た。また、静止画像に対して、カテゴリの分類と位置情報は正しく表示された。ただ、動画の認識率はあまり良くない。動画の認識率は 5NM 未満で約 70%、5NM と 10NM の間で約 60%、10NM 以外約 40%である。太陽光や天候、波、島などの影響がある同時にデータセットの分類精度を改善する必要があると思う。動画の認識結果を見ると、航海環境が認識結果に大きな影響を与えたと考える。天気が良くない場合には、視野も不良になる。例えば、晴の日に認識率は霧の日より高い。また、学習モデルはカメラで撮影された画像を学習する。しかし、実験の時には汐路丸に設置した IP カメラからの航海映像の認識実験を行う。それも認識結果に影響を与える。

動画を用いた実験の認識精度の向上のために、航海映像を用いた約 5,700 枚の学習用データセットを作った。海平線の近くの障害物を5つのカテゴリ(11(近いところの大型船)、12(遠いところの大型船)、s1(近いところの小型船)、s2(遠いところの小型船)、buoy)に分類した。そのデータセットを約 200epoch 訓練し、学習モデルをこの weight でかけて障害物を認識する。航行映像に対して認識率は上がったが、理想な結果ではない。航行映像を用いた実験ではまた未検出と誤検出の場合がある。

LSTM を用いた実験では静止画像の認識は悪い結果を得た。枠は正しくない。また、LSTM の船舶認識も行ったが、何にも検出していない。映像の認識はすごく遅く、毎秒 1 FPS になっている。

11、考察

SSDにより障害物検出は良い結果も出たが、小型船やブイなどの小物体はほとんど検出できなかった。航海環境はある程度上に認識結果に影響を与える。航海実験を行うとき、晴の日とは限らない。霧や雨の場合も多い。視野不良の時に、認識率は高くなく、未検出と誤検出が多い。航海環境の原因を除いて、ある程度にSSDネットワーク構造と関係がある。SSDでは、入力画像がその大きさに関わらず、読み込むときに、自動的に300*300ピクセル画像へ変換する。その特性により、1980*1980ピクセル元画像中の小型船やブイなどの小物体は消える可能性がある。モデルが学習するときにも学習できない。そして、未検出と誤検出の原因は海上環境の影響だけではなく、データセットもひとつの原因となっている。また、汐路丸は他の船舶に近づくと、衝突の危険がある。そのため、撮影された船舶が小さく、検出に影響を及ぼす。汐路丸と物標との距離により、固定カメラを利用する現状の環境では小型の物体の認識が困難である。それもひとつの原因となっている。

LSTMに関して、私たちは学習済みのモデルを利用し、映像の船舶認識を行ったが、何にも検出していない。また、映像の表示はすごく遅く、毎秒1FPSになっている。同一物体を連続検出に向けてLSTMを用いた船舶認識を試みたが、必要なデータが保持されていない。また、動画認識の方はGPUのメモリ容量が足りないと考えている。なぜならば、SSDのフレームワークtensorflowは計算量が膨大なため、システム実行の効率が低下し、メモリ容量の使用量が大きい。LSTMを用いたSSDのネットワークの構造に関してまた検討する必要がある。そのため、障害物検出の向上と複数船舶検出手法の検討、船舶認識結果に影響する原因を検討する必要がある。また、時系列データとSSDネットワーク構造の矛盾を解決する必要がある。

12、まとめ

本研究では、SSD により、障害物を 9 つのカテゴリを分類し、学習モデルを作成した。最優な 50epoch 重みの学習モデルを用いて静止画像と海上航行映像に関する実験を行った。また、航海映像からの障害物がすごく小さく、検出が難しいこの問題を解決するために、海平線の近くの障害物を 5 つのカテゴリを分類し、学習モデルを作成した。最優な 200epoch 重みの学習モデルを用いて海上航行映像に関する実験を行った。本研究で得られた成果を以下に示す。

- (1) 海上航行映像からコンテナ船、LNG 船、PCC、その他の大型船、漁船、プレジャーボート、タグボート、その他の小型船、ブイの 9 のカテゴリのデータセットを作成した。
- (2) 作成したデータセットを利用し、海上を航行する船舶にとって障害となる障害物の検出と表示を行い、評価を行った。
- (3) 海平線の近くの障害物を 5 つのカテゴリ、l1(近いところの大型船)、l2 (遠いところの大型船)、s1 (近いところの小型船)、s2 (遠いところの小型船)、buoy に分類した。その五つカテゴリのデータセットを作成した。
- (4) この 2 種類のデータセットを SSD で学習し、それぞれの最優な重みを取得した。
- (5) それらの重みを SSD 予測モデルにかけ、海上航行中の他船やブイなどの障害物に関する認識、表示する実験を行った。
- (6) 本研究では、海難事故削減ために、機械学習を用いた船舶認識手法を提案した。
- (7) 障害物認識精度向上に向け、時系列データを用いた船舶 認識モデルを検討した。

静止画像に対しては良い 結果を得たが、動画像に対しては認識率の向上が必要である。時系列を用いた予測では、 LSTM の追加場所/情報などモデル構成から再検討する必要がある。

今後の課題としては、障害物検出の向上と複数船舶検出手法の検討、船舶認識結果に影響する原因、航路予測手法を検討する。また、学習回数数と検出率の関係、時系列データと検出率の関係を検証する。高精度の物体検出手法の提案と船舶の自動運転の実現を進める。

13、参考文献

- [1]海上保安庁：平成 29 年海難の現況と対策
- [2]海上衝突予防法
- [3] SW/HW 協調処理による船舶自動見張りシステム-船舶発見に用いる画像認識の高速化, 高橋良平,植竹大地,大川猛,横田隆史,大津金光,松本洋平,2014.
- [4] AIS 情報を用いた小形船用見張り支援システム, 丹羽 康之,福戸 淳司,宮本 佳則,稲岡 孝,海上技術安全研究所報告, pp107, 2010 年 3 月.
- [5] 小型船舶の狭水路航行支援に関する研究－河川交通に向けて－, 羽子田 哲弥, 大石 浩世, 古谷 雅理, 竹本 孝弘, 日本航海学会論文集 [0388-7405], pp88-97, 2017
- [6] 見張り支援ニューラルネットワークに関する基礎研究, 稲石 正明, 足立 浩一, 山下 千春, 杉崎 昭生, 日本航海学会論文集[0388-7405], pp203-211, 1994.
- [7] 航海環境画像を用いた障害物分類に関する研究 -河川交通に向けて-, 近藤 正樹, 稲石 正明, 古谷 雅理, 中川 正樹, 日本航海学会論文集[0388-7405], pp144-151, 2017.
- [8] Rich feature hierarchies for accurate object detection and semantic segmentation,Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik(Submitted on 11 Nov 2013 (v1), last revised 22 Oct 2014 (this version, v5)) arXiv:1311.2524 [cs.CV]
- [9] Selective Search for Object Recognition, J.R.R.Uijlings, K.E.A. van de Sande, T. Gevers, and A.W.M. Smeulders,Technical Report 2012, submitted to IJCV
- [10] Mask-CNN: Localizing Parts and Selecting Descriptors for Fine-Grained Image Recognition,Xiu-Shen Wei, Chen-Wei Xie, Jianxin Wu(Submitted on 23 May 2016), arXiv:1605.06878 [cs.CV]
- [11] 入門 機械学習,Drew Conway , John Myles White (著), 萩原 正人, 奥野 陽 , 水野 貴明 , 木下 哲也 (訳),2012.
- [12] IT Text 人工知能 改訂 2 版, 松本 一教, 宮原 哲浩, 永井 保夫, 市瀬 龍太郎, 本位田 真一, 第五章第九節,pp141-158,2015.
- [13] 入門 機械学習,Drew Conway , John Myles White (著), 萩原 正人, 奥野 陽 , 水野 貴明 , 木下 哲也 (訳),第十二章,pp 293-304,2012.
- [14] 入門 機械学習,Drew Conway , John Myles White (著), 萩原 正人, 奥野 陽 , 水野 貴明 , 木下 哲也 (訳),第五章,pp 135-163,2012.
- [15] IT Text 人工知能 改訂 2 版, 松本 一教, 宮原 哲浩, 永井 保夫, 市瀬 龍太郎, 本位田 真一, 第五章第七節 pp128-138,2015.
- [16] Machine Learning, Leo Breiman,pp 5–32,Volume 45, Issue 1, October 2001.
- [17] 入門 機械学習,Drew Conway , John Myles White (著), 萩原 正人, 奥野 陽 , 水野 貴明 , 木下 哲也 (訳),第十章,pp 247-257,2012.
- [18] 入門 機械学習,Drew Conway , John Myles White (著), 萩原 正人, 奥野 陽 , 水野 貴

- 明, 木下 哲也 (訳),第六章第三節,pp 186-193,2012.
- [19] R. Agrawal, T. Imielinski, and A. Swami(1993). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pages 207-216.
- [20] IT Text 人工知能 改訂 2 版, 松本 一教, 宮原 哲浩, 永井 保夫, 市瀬 龍太郎, 本位田 真一,第五章第八節,pp139-140,2015.
- [21] Howard, Ronald. A. (1960). Dynamic Programming and Markov Processes. The M.I.T. Press.
- [22] Discriminant Adaptive Nearest Neighbor Classification,Trevor Hastie,Rolbert Tibshirani,IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 18, NO. 6, JUNE 1996
- [23] Nomograms for Visualization of Naive Bayesian Classifier, Mozina M, Demsar J, Kattan M, Zupan B, pp337-348,In Proc. of PKDD-2004,2004.
- [24] V. Vapnik and A. Lerner. Pattern recognition using generalized portrait method. Automation and Remote Control, 24, 1963.
- [25] IT Text 人工知能 改訂 2 版, 松本 一教, 宮原 哲浩, 永井 保夫, 市瀬 龍太郎, 本位田 真一, 第一章第二節 pp 3 ,2015.
- [26] A Critical Review of Recurrent Neural Networks for Sequence Learning,Zachary C. Lipton, John Berkowitz, Charles Elkan(Submitted on 29 May 2015 (v1), last revised 17 Oct 2015 (this version, v4)), arXiv:1506.00019 [cs.LG].
- [27] 深層学習 Deep Learning (監修:人工知能学会),麻生 英樹, 安田 宗樹, 前田 新一, 岡野原 大輔, 岡谷 貴之, 久保 陽太郎, ボレガラ ダヌシカ (著), 人工知能学会 (監修), 神 鷲 敏弘 (編集),pp153-165,2015.
- [28] Rich feature hierarchies for accurate object detection and semantic segmentation,Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik(Submitted on 11 Nov 2013 (v1), last revised 22 Oct 2014 (this version, v5)), arXiv:1311.2524 [cs.CV],2013.
- [29] Fast R-CNN,Ross Girshick(Submitted on 30 Apr 2015 (v1), last revised 27 Sep 2015 (this version, v2)), arXiv:1504.08083 [cs.CV],2015.
- [30] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, (Submitted on 4 Jun 2015 (v1), last revised 6 Jan 2016 (this version, v3)),arXiv:1506.01497 [cs.CV]
- [31] Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, Kaiming He ,Xiangyu Zhang ,Shaoqing Ren,Jian Sun, Volume: 37, Issue: 9, Sept. 1 2015.
- [32] SSD: Single Shot MultiBox Detector, Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg,(Submitted on 8 Dec 2015 (v1), last revised 29 Dec 2016 (this version, v5)),arXiv:1512.02325 [cs.CV]

- [33] You Only Look Once: Unified, Real-Time Object Detection, Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi (Submitted on 8 Jun 2015 (v1), last revised 9 May 2016 (this version, v5)), arXiv:1506.02640 [cs.CV], 2015.
- [34] Selective Search for object recognition, J.R.R. Uijlings University of Trento, Italy, IJCV 2012, 2012.
- [35] 深層学習 Deep Learning (監修:人工知能学会), 麻生 英樹, 安田 宗樹, 前田 新一, 岡野原 大輔, 岡谷 貴之, 久保 陽太郎, ボレガラ ダヌシカ (著), 人工知能学会 (監修), 神 島 敏弘 (編集), pp211-214, 2015.
- [36] Speech recognition with deep recurrent neural networks, Alex Graves, Abdel-rahman Mohamed, Geoffrey Hinton
- [37] A Critical Review of Recurrent Neural Networks for Sequence Learning, Zachary C. Lipton, John Berkowitz, Charles Elkan.
- [38] Bell S, Zitnick C L, Bala K, et al. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. In CVPR 2016.
- [39] CNN-RNN: A Unified Framework for Multi-label Image Classification, Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, Wei Xu, arXiv:1604.04573 [cs.CV]
- [40] 深層学習 Deep Learning (監修:人工知能学会), 麻生 英樹, 安田 宗樹, 前田 新一, 岡野原 大輔, 岡谷 貴之, 久保 陽太郎, ボレガラ ダヌシカ (著), 人工知能学会 (監修), 神 島 敏弘 (編集), pp214-217, 2015.
- [41] Gers, Felix A., and J. Schmidhuber. "Recurrent Nets that Time and Count." Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00)-Volume 3 - Volume 3 IEEE Computer Society, 2000:3189.
- [42] Semantic Object Parsing with Graph LSTM, Xiaodan Liang, Xiaohui Shen, Jiashi Feng, Liang Lin, Shuicheng Yan, arXiv:1603.07063 [cs.CV]
- [43] Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, Yoshua Bengio (Submitted on 11 Dec 2014), arXiv:1412.3555 [cs.NE], 2014.
- [44] Quasi-Recurrent Neural Networks, James Bradbury, Stephen Merity, Caiming Xiong, Richard Socher, Nov 2016.

謝辞

本研究を遂行し修士論文をまとめるに当たり、多くのご支援とご指導を賜りました、指導教員である古谷雅理准教授に深く感謝しております。古谷先生は忙しいですが、しばしば私たちに研究の進捗状況を聞き、辛抱強くご指導を賜りました。論文を書く際には、熱心なご指導をいただき、多くの有用な情報を提供します。まだまだ至らない点があった私の研究へ丁寧なご指導ご鞭撻を頂いたことに、心より感謝申し上げます。また、研究室の近藤正樹技術職員はお忙しい中、私の研究丁寧なご助言とご指導を賜ったこと、感謝申し上げます。また、データセットの作成のご協力をいただいたことに、感謝の意を表します。

私の未熟の研究に対して常に厳しくも温かいご指導をしてくださった稲石正明教授に感謝申し上げます。機械学習の研究が難しいが、稲石先生は終始優しい言葉で私を励まして下さいました。私の研究にコメントを下さりまして、本当にありがとうございました。また、お忙しい中であるのに関わらず、時間を割いて査読をお引き受けいただきました庄司るり教授に心よりお礼申し上げます。

その後、同じ研究室の閻氏、三宅氏また同じ研究室のメンバーに感謝いたします。閻さんと一緒に船舶認識を研究する過程に、参考になるお話を聞かせていただくことができました。また、研究がうまくいかなかったときにも励まし合っています。研究以外に何度かご自宅に招待していただくなど、生活上にもいろいろなお世話になりました。三宅さんは自分の研究ではないにもかかわらず、汐路丸実験で数回も乗船し、私の実験に大きな助けをいただきました。心より感謝します。また、同じ研究室の保坂さんと安原さんは学習用データセットの作成にご協力をいただき、誠にありがとうございました。みんな個性であり、おかげで楽しく過ごすことができました。研究室の内外で一緒に行動する機会が多く、学生時代の良い思い出を作ることができました。また、学習環境と学業支援を提供してくださった東京海洋大学にお礼を申し上げます。学校の益々の発展を心よりお祈り申し上げます。最後に、これまで自分を育ててくれた両親に感謝致します。家庭負担が重いですが、私の海外留学を支援しています。生活上に経済的支援を与えてくれますし、学業上に精神的サポートを与えてくれます。本当にありがとうございました。

留学の2年間に、いろいろな知識と経験を積みました。これらの知識と経験は私今後の人生に一生適用できます。機械学習の研究は最初にうまく進んでいなく、実験も何度か失敗したことがありました。しかし、古谷先生は私に温かく励まして下さった同時に、参考論文などの有用な情報を提供します。2年間の研究を通じ、知識の獲得だけではなく、堅持の大切さを学びました。これから、社会人になり、今まで積んだ知識と経験を活かし、自分自身の価値を実現する同時に、社会に貢献したいと考えております。