

物体検出とセマンティックセグメンテーションによる 車道上に限定した未知物体検出手法の提案

別宮 広朗^{1,a)} 河崎隆文^{2,b)} 中澤 仁^{1,c)}

概要：現在、道路上の廃棄物は、国や県といった道路の管理責任者に報告することで除去される。道路の利用者に報告を大きく委ねているため、長期的に廃棄物が放置される可能性があり、責任者が日常的に街全体の調査をするのは費用が莫大になってしまう。本研究では、道路上の異状の中でも、車道上の廃棄物に注目し、車道上の未知の物体を自動で検出するシステムの構築手法を提案する。交通に影響のある車道上の廃棄物などを自動で検出後、通知するシステムを構築し、町全体を日常的に徘徊するゴミ収集車などに設置することで、廃棄物の長期的な放置されることを防ぎ、調査する費用を削減することが可能である。本研究の手法は複数の機能によって実現する。一つ目の機能は、物体と思われるものを全て検出するものであり、学習データに存在しない未知の物体に対しても検出することを可能にする。二つ目の機能は、車道上に限定した物体検出を行うために、セマンティックセグメンテーション技術を用いて、全ての物体を無視して車道だと思われる領域を予測する。物体を検出し、予測された車道領域に含まれない物体を取り除くことで、車道上にのみ存在する物体の検出を可能とする。上記の提案手法は、領域に限定した未知の物体のデータセットの作成の困難さを回避しているため、容易に実装が可能である。本研究では、各機能と提案手法の性能の評価を独自に作成したデータセットなどを用いて行い、各機能と提案手法の有効性を示す。

1. 序論

道路上には、日常的に異状が存在する可能性があり、道路の穴や廃棄物、道路の汚れなど様々である。管理責任者が日常的に街全体の道路を調査するには費用が莫大になってしまうため、国や県といった道路の管理責任者に報告することで除去されるのが一般的である。しかし、報告がない場合、管理責任者が道路の異状に対応することができないため、道路の利用者に解決を大きく委ねることになる。そのため、道路上の異状を検知して、自動で道路管理責任者に通知するシステムがあれば、上記の問題が解決される。[1]では、物体検出技術を用いて、画像から道路の損傷検出を行っており、車載カメラや監視カメラによって道路を監視して、特定の道路上の損傷を検出する。将来的に、町を徘徊するゴミ収集車やバスなどに取り付けることで、自動で町全体の道路の損傷を検出するシステムを作成することが可

能である。我々の研究は、道路異状の中で車道上の廃棄物に注目しており、上述した構想と同様に、通知システムを徘徊するゴミ収集車やバスなどに取り付けることで、廃棄物が長期的に放置されることを防ぎ、町全体を調査する費用を削減することが可能である。

本研究では、車道上の廃棄物などの未知の物体を自動で検出するシステムの構築手法を提案する。提案する手法は、複数の機能に分解することができる。廃棄物は、定めることのできない多様な物体であり、事前に全てを網羅する廃棄物を学習データとして用意するのは困難である。そのため、一つ目の機能は、物体と思われるものを全て検出することである。また、歩道や車道以外の場所に廃棄物や未知の物体が存在することは許容する必要があるため、二つ目の機能は、車道上に限定した物体検出を行うことである。本研究の目的は、未知の物体に対して注目しているため、既知物体を検出対象から取り除く必要がある。そのため、三つ目の機能は、検出された車道上の物体に対して、車、人、自転車、バス、トラックなどの既知の車道に存在していても問題のない物体を分類して取り除くことである。

一つ目の機能は、物体らしさを検出することであり、既知の物体に対するラベルによって学習される従来の物体検出とは目的が異なる。本研究では、学習データに存在しない未知の物体に対しても検出することを可能にする手法と

¹ 慶應義塾大学環境情報学部
Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa 252-0882, Japan

² 慶應義塾大学大学院政策・メディア研究科
Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa 252-0882, Japan

a) t18727hb@sfc.keio.ac.jp

b) drgnman@ht.sfc.keio.ac.jp

c) jin@ht.sfc.keio.ac.jp

して、PascalVOC データセットのラベルを全て一つのオブジェクトとして学習することで実現する。

二つ目の機能は、領域に限定した物体を検出することであり、提案する上で主に二つの手法が考えられる。車道上の物体と車道外の物体を分類して検出する手法と、車道をセマンティックセグメンテーションによって領域を検出し、車道上に存在する物体のみを検出対象とする手法である。前者に比べると後者は、処理が多く非効率なシステム構造であるといえる。しかし、前者は、車道上の物体と車道外の物体のデータをそれぞれ用意する必要がある。更に、その物体は未知の物体検出を可能にするために多様なものがある必要がある。このようなデータセットを作成することは極めて困難であり、道路や車道に限ったデータセットではない PascalVOC などのデータセットの適用もできない。未知の検出領域を広げるために新しい物体の画像データを用意する上でも、車道上と車道外の画像データを用意するため、柔軟性も低い。そのため、本研究では、後者の手法を採用する。セマンティックセグメンテーション技術を用いて、全ての物体を無視して車道だと思われる領域を予測する。物体を検出し、予測された車道領域に含まれない物体を取り除くことで、車道上にのみ存在する物体の検出を可能とする。

三つ目の機能は、一つ目の機能で検出された物体から既知物体を検出対象から排除することである。本研究では、物体検出モデルを用いて既知か未知かどうかを分類する。この処理に関しては、分類するためのデータセットを作成することで、画像分類問題として扱え、物体検出モデルよりも軽量な分類モデルを適用することが可能である。

上記の提案手法の検出性能を、独自に作成した車道上の未知の物体のデータセットで評価することで、手法の有効性を確認する。本研究の貢献は以下の通りである。

- 領域上における物体のデータセットが存在しない場合でも、領域上の物体検出を実現するための手法を提案した。
- 提案した手法の各機能である、物体らしさの検出と車道を検出するためのセマンティックセグメンテーションの有効性を検証した。また、物体らしさの検出に関しては、考察において提案手法のモデルが通常の学習モデルよりも適していることを示す。
- 車道上に限定した未知の物体を検出するための提案手法の有効性を独自によって作成したデータセットによって検証した。

本論文の構成は、2 章では物体検出や未知の物体の検出などの関連研究を述べ、3 章では、提案手法の各機能における詳細な手法を解説する。4 章では、提案手法の各機能における性能の評価を様々なデータセットを用いて行う。5 章では、実験結果や実験によって得た知見に対して考察し、6

章では、提案手法の課題や展望をまとめ、7 章では、本研究の結論を述べる。

2. 関連研究

物体検出は画像を入力として受け取り、数種類の学習された物体に対して、バウンディングボックスと呼ばれる物体を四角く囲うような座標を出力する技術であり、様々な分野で活用されている。特に本研究の目的に類似する、道路上の異状を検出するために物体検出を応用している研究も少なくない。[1][2] では、道路上の損傷画像のデータセットを作成し学習された物体検出モデルに対して、スマートフォン画像を入力とし、道路損傷検出と分類を行なっている。これらの研究は、車道上のような領域に限定した物体検出を行えているわけではないため、本研究の目的に対する手法としてそのまま応用することはできない。

ある領域に限定した物体検出の研究では [3] があり、モバイル環境の運用コストの削減を目的に、道路上における危険な注意すべき領域をクラス活性化マッピングを通して取得し、その領域にのみ物体検出を行なっている。複数のモデル構造を要することなく出力できているが、注意領域の出力には、既知の物体に定義した危険度というような概念によって領域が抽出されているため、領域が物体と関連性の低い場合や、物体そのものが未知である時、別の領域推定機構が必要となる。

未知の物体検出は非常に難しい課題であり、既知のクラスと未知のクラスの両方の物体を検出できることを目的とすることをオープンセット認識と呼ばれる。[4] では、物体の確信度による閾値を学習して、閾値以下の確信度の物体を未知の物体と分類することで、既知と未知の物体を同時に検出することを実現させている。[5] では、検出した物体の分類に成功した時に限りペナルティを与えることで、学習データのクラスへの依存を軽減し、学習データに存在しない物体らしさを学習することを目標としている。結果的に、未知の物体の検出の精度を向上させ、手法の有効性を示している。検出以外の研究では [6] があり、未知の物体と既知の物体のラベル空間に対して、敵対的学習を用いて分類を可能にし、ドメインアダプテーションの分野に適用している。[7] では、LiDAR 点群のための既知のクラスと未知のクラスの両方の物体に対して、深層学習によって不確かな点を識別することで、オープンセット認識を可能としている。

3. 提案手法

3.1 物体らしさの検出

本研究での物体検出は、学習データのラベル上に存在しないような未知の物体に対しても検出を行うことが目標である。そのため、PascalVOC データセットの様々なラベルを一つのオブジェクトとしてラベル付けした学習データを

利用する。これは、既知の物体から、全ての物体に通じるような物体らしさを広く学習することを目的としており、特徴量が分類される様々なラベルの領域を一つのオブジェクトというラベル領域に広げることで、従来では属さなかった各ラベルの中間の空間に対しても、オブジェクトという広いラベル領域に属すことを期待している。

本研究では、物体らしさの検出モデルは CenterNet[8] であり、バックボーンモデルは 101 層の ResNet[9] である。CenterNet は、最新のモデルではなく、最先端の性能とはいえない。物体らしさの検出においては、高性能のモデルでなければ機能しないということではなく、提案手法を検討する上でのベースラインとして CenterNet を利用している。

3.2 セマンティックセグメンテーションによる車道領域の検出

本研究では、物体検出を行いたい領域は、車道上に限定している。通常の物体検出モデルは、領域に拘っていないため、ある領域上に限定した検出は不可能である。画像におけるピクセル毎の領域の分類は、出力をピクセル毎の分類問題として扱うセマンティックセグメンテーションモデルによって実現できる。そのため、物体検出とセマンティックセグメンテーションの二つの技術を用いることで、領域上の物体検出が可能そうである。

物体検出モデルと併用する場合、このセマンティックセグメンテーションモデルは物体を認識する必要はない。車道上に物体が存在していた場合、画像上では車道の上に被さるように表示されるため、セグメンテーションされる物体の画像部は車道外と出力される。その場合、検出される物体のバウンディングボックスの座標と同じ座標のセグメンテーションモデルの出力を確認しても、物体の位置が車道外と判断されてしまう。期待される機能は、物体が被さる位置が本来車道であるかどうかの判別であるため、ここでのセマンティックセグメンテーションは、物体を無視して車道か車道外を判別する機構である。そのため、このセマンティックセグメンテーションモデルは、図 1 のように物体を無視して、車道らしい場所をアノテーションしたデータセットによって学習される。

本研究での、セマンティックセグメンテーションモデルは、segmentation_models_pytorch の UNet++[10] であり、エンコーダー部分のモデルは efficientnet-b7[11] である。ここでのセマンティックセグメンテーションにおいても、物体らしさの検出と同様に、機能する上で高性能なモデルである必要はないため、提案手法を検討する上でのベースラインとして Unet++ を利用している。

3.3 領域上の物体の検出

3.1 と 3.2 の技術を用いることで、限られた領域上の物体の検出を行える。3.1 の物体検出によって得られるバウン



図 1 物体を無視して車道らしい場所をアノテーションした、本研究におけるセマンティックセグメンテーションモデルの学習用の画像である。

ディングボックスの座標と車道だと思われるセグメンテーションの座標を比較して、図 2 のように車道領域であれば検出対象とし、車道領域でない場合は、検出対象から取り除く。比較する座標は、バウンディングボックスの右下と左下の座標である。これは、座標の右上と左上の場合は、吊るされた旗や信号機などの座標が車道領域と被さっていたとしても車道領域に存在していない場合が考えられるためである。

物体を無視せずに車道と車道外の領域のセグメンテーションを行えば、物体が写る画像ピクセルは車道外として分類されるため、物体検出を用いずにセグメンテーションのみで検出することが可能そうである。しかし、この検出の場合、物体であることではなく車道外であるということに注目して検出しているため、車道外に分類されたものが物体としての確証を得られるわけではない。更に複雑な車道の形状の場合、車道上に密接した車道外の部分を物体として捉えかねない。また、車道と車道外、物体の 3 種類のアノテーションされたデータセットによって学習することでセグメンテーションモデルのみで未知の物体の検出も可能そうである。しかし、物体らしさを検出する上では多様な物体のデータセットから学習する必要がある一方で、自然に車道上に存在する物体というのは限られる。そのため、PascalVOC データセットのような多様な物体を車道上に配置した画像を取得し、アノテーションを行う必要があり、データセットの作成が非常に困難である。よって本研究手法では、車道上に限定した物体検出のために、物体検出とセグメンテーションの併用をしている。

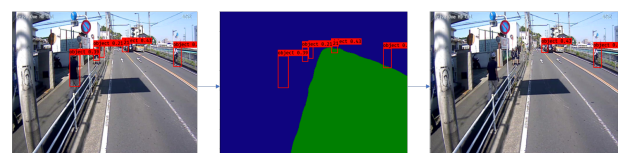


図 2 左図が物体検出の画像であり、中央図が座標比較のイメージ図である。右図は車道領域外に存在する物体を取り除いている。

3.4 既知物体の検出の除外

本研究の目的は、車道上の未知の物体の検出であり、3.1

の物体検出モデルでは、物体らしい物体を総じて検出してしまいうため、更に改良が必要である。我々はそこに検出された物体の画像を切り取り新たな画像とし、学習済みの高精度な物体検出モデルを経由し分類することで、3.3の検出対象から車や人、自転車などの既知の物体を取り除く。この時、確信度が0.5以上の物体のみを取り除いており、低い確信度の場合は既知であるとは言えないとしている。ここでは、必ずしも物体検出モデルである必要はないが、性能が低いと3.3の出力に悪影響を与えてしまうため、高性能である必要がある。人や車、自転車などを含むデータセットの学習済みの高精度なモデルとして、COCO データセットによって事前に学習された Yolov5[12] を利用している。

3.5 提案手法の全体の構造

画像を入力として受け取り、CenterNet によって物体らしいものを検出し、Unet++によって画像の車道と車道外の領域をセグメンテーションする。二つのモデルの出力結果から車道領域外に存在する物体は検出対象から除外し、既知物体検出器として機能する Yolov5 によって物体を分類し既知の物体は取り除くことで、車道上に限定した未知の物の検出を実現することが本研究で提案する手法の全体の構造である。図3は、本研究の提案手法の全体の構造を表したものである。

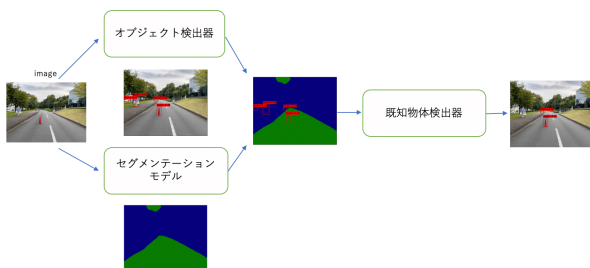


図3 車道上に限定した未知の物体検出を目的とした、本研究の提案手法の全体像である。

4. 実験

本章では、提案手法の有効性を実験を通して、検証する。

4.1では、物体らしさの検出の性能を評価する。テストに用いるデータセットは、ラベルをオブジェクトに統一した COCO データセットを利用する。学習データとして用いた PascalVOC データセットのクラス数は20であり、COCO データセットのクラス数は80である。person や car のような PascalVOC データセットと COCO データセットの両方に含まれるクラスも存在するが、COCO データセットの方が検出対象のオブジェクトクラスが多様である。そのため、PascalVOC データセットのみで物体らしさを検出するモデルが多様なオブジェクトを持つ COCO データセットに対する検出性能から、どの程度物体らしさを検出できて

いるかを評価する。また、PascalVOC データセットに存在しないクラスのみ注目した COCO データセットを用いて、未知の物体における検出性能を評価する。COCO データセットをテストデータとする場合、未知のみならず既知の物体の検出性能も評価するため、性能評価の評価指標は、mAP を用いる。既知の物体の検出性能を評価する理由は、提案手法が全ての物体を一通り検出することを想定しているためである。PascalVOC データセットに存在しないクラスのみ注目した COCO データセットをテストデータとする場合は、未知の物体をどの程度検出しているかを評価するため、性能評価の評価指標は Recall を用いる。未知の物体の検出は、ラベルを超えた物体を検出するため、COCO データセットの正解ラベルによって mAP で評価してしまうと存在しないクラスの物体の出力は偽陽性と判断されて評価が低下してしまう。そのため、偽陽性に寛容な Recall で評価を行う。

4.2では、物体を無視した車道と車道外の領域のセマンティックセグメンテーションの性能を確認する。データセットは、車載カメラの動画を独自にアノテーションしたものであり、329枚の訓練データから学習され、141枚のテストデータによって検証される。性能評価の評価指標は、F値と IOU スコアを用いる。

4.3では、それぞれ個別に学習されたモデルを併用し、3.5で紹介したような提案手法による、車道上に限定した未知の物体の検出性能を評価する。テストデータは、独自に作成した45枚の車道や車道外の領域に物体が存在するデータセットを用いる。性能評価の評価指標は、mAP を用いる。

4.1 物体らしさの検出の性能評価

PascalVOC データセットのラベルをオブジェクトに統一して学習したモデルの COCO データセットのテストデータに対する性能を表1に示す。PascalVOC データセットにはアノテーションされていないような物体を多数含む COCO データセットの物体を検出することは困難であるにも関わらず、IOU の閾値を0.5の時の mAP は0.722であり、異なるデータセットに対しても物体をある程度検出できていると言える。本研究の実験では、物体検出のモデルとして CenterNet を使用しているが、性能の良いモデルにすることで mAP の値は向上していくと考える。

検出モデル	$mAP_{0.25}$	$mAP_{0.5}$	$mAP_{0.75}$
CenterNet	0.786	0.722	0.538

また、PascalVOC データセットに存在しないクラスのみ注目した COCO データセットのテストデータに対する、性能を表2に示す。IOU の閾値を0.5の時の Recall の値は、

0.104 であり、通常の学習では見過ごされるようなクラスの物体を検出することが可能となっていることがわかる。

表 2 PascalVOC データセットに存在しないクラスのみ注目した COCO データセットのテストデータによる、未知の物体の検出の性能

検出モデル	$R_{0.25}$	$R_{0.5}$	$R_{0.75}$
CenterNet	0.146	0.104	0.058

4.2 物体を無視した車道と車道外の領域のセマンティックセグメンテーションの性能評価

独自にアノテーションした 141 枚のテストデータに対する、物体を無視した車道と車道外を分類するセマンティックセグメンテーションモデルの性能を表 3 に示す。F 値は 0.976, IOU スコアは 0.964 であり、非常に高い精度でセグメンテーションができていていると言える。図 4 は実際に、出力される画像の一例であり、物体を無視した車道と車道外の領域のセグメンテーションが高い精度で実現している。

表 3 車道と車道外の領域のセマンティックセグメンテーションの性能

検出モデル	F 値	IOU スコア
Unet++	0.976	0.964

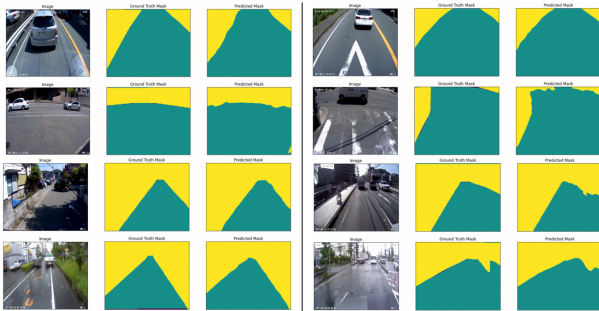


図 4 テストデータに対するセグメンテーションの実行例である。左図が入力画像であり、中央図がアノテーション画像であり、右図が Unet++ による予測画像である。

4.3 提案手法による、車道上の未知の物体検出の性能評価

独自にアノテーションした 45 枚のテストデータに対する、車道上に限定した未知の物体検出の性能を表 4 に示す。テストデータに含まれる物体は、人、ペットボトル、木片、リュックサック、靴、バケツ、三角コーン、コーンバーなどが含まれる。閾値が 0.5 程度までは、mAP は 0.7 を保っているものの、閾値が 0.75 を超える場合、mAP は 0.223 と性能が大きく下がっている。

図 5 は、車道上に限定した未知の物体検出の実験に用い

表 4 車道上に限定した未知の物体検出の性能

	$mAP_{0.25}$	$mAP_{0.5}$	$mAP_{0.75}$
提案手法	0.780	0.720	0.223

たテストデータに対する、提案手法による、検出結果の一例である。

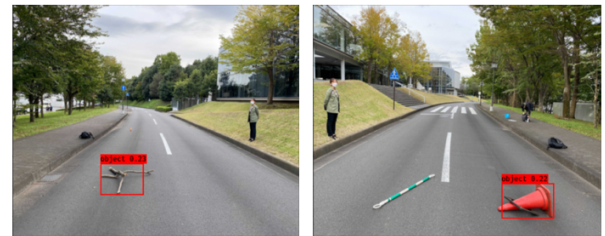


図 5 4.3 の実験の提案手法による、検出結果の一例である。

5. 実験や結果による考察

実験を通して得た知見とともに考察をまとめる。

5.1 実験結果からの考察

PascalVOC データセットに存在しないクラスの COCO データセットをテストデータとしたとき、IOU の閾値を 0.5 とした Recall の値が 0.104 というのは、一見低く見える。これは、COCO データセットの難易度によるものなのか、物体らしさを検出する能力を得ていないのか判断する必要がある。COCO データセットの内、PascalVOC データセットに存在するクラスのみを対象としてテストした時の性能を評価したものが表 5 である。既知のクラスの COCO データセットをテストデータとしても、IOU の閾値を 0.5 とした Recall の値は 0.273 と低く、COCO データセットの物体を検出する難しさがわかる。未知の物体の検出性能は、既知の物体の約 4 割の評価値を記録しており、物体らしさを検出する能力は十分習得していると言える。

表 5 PascalVOC データセットに存在するクラスのみ注目した COCO データセットのテストデータによる、物体の検出の性能

検出モデル	$R_{0.25}$	$R_{0.5}$	$R_{0.75}$
CenterNet	0.322	0.273	0.169

物体を無視したセマンティックセグメンテーションは非常に高い精度を実現している。これは、物体が存在し、画像上で背景に被さっていたとしても、背景がどんな領域であるかの推定を十分行えることを示しており、提案した学習手法は、様々な分野へと応用できる可能性がある。

図 5 からわかるように、セマンティックセグメンテーションによる領域の推論の性能が良いため、車道上に限った物体検出は高い精度で行えている。検出としての誤りは、物

体らしさの検出の困難さから、未知の物体を見過ごす場合が殆どである。一方で、偽陽性によつての誤りは、偽陰性に比べて非常に少ないため、検出による見過ごしを軽減するように、モデルの学習過程で調節することで性能の向上が期待できそうである。

5.2 性能の低い通常の物体検出モデルによる、未知の物体の検出の可能性

本研究の提案手法は、最後に既知物体かどうかの分類するモデルを組み込むことで、車道上に存在しても問題ない車や人を検出対象から取り除いている。これは、最初の物体らしいものを検出する物体検出モデルが一つのオブジェクトとしてしか検出することができないことが原因である。物体らしいものを検出することが、通常の物体検出モデルで可能であるならば、確信度が高い人や車を取り除けば良いだけなので、既知物体かどうかを検出するモデルが必要なくなる。PascalVOC データセットを 20 種類のラベルを変更することなく学習した、物体らしさを検出した同モデルの CenterNet を用いて、4.1 と同様の実験を行なった結果が表 6 である。今回の場合、ラベルをオブジェクトに統一した COCO データセットをテストデータにした場合、通常の物体検出モデルの方が性能が良い。これは、本研究の提案手法にとっては残念な事実であるが、COCO データセットと PascalVOC データセットは人や車の物体が大量であり、PascalVOC データセットを適用させやすいような類似するデータセットであることが要因だと予想される。そのため、未知のみの物体のデータセットを用いて、性能を正當に評価する必要がある。

表 6 通常の物体検出モデルとラベルをオブジェクトに統一して学習したモデルによる、物体らしいものの検出の性能比較

検出モデル	$mAP_{0.25}$	$mAP_{0.5}$	$mAP_{0.75}$
通常 CenterNet	0.807	0.755	0.566
提案 CenterNet	0.786	0.722	0.538

4.1 の実験でも未知の物体の検出性能を評価するために使用した、PascalVOC データセットに存在しないクラスのみ注目した COCO データセットのテストデータを用いる。未知の物体の検出性能を比較したものが表 7 である。未知の物体の検出の性能では、ラベルをオブジェクトに統一したモデルが全ての閾値において優れた。未知の物体のデータセットでも性能が通常の物体検出モデルの方が性能が良い場合、提案手法でも通常の物体検出モデルを採用するが、本研究では未知の物体の検出を目的としているため、未知の物体のデータセットに対して優れた、ラベルをオブジェクトに統一したモデルを提案手法として採用した。

また、他の例として学習時のクラスに含まれない未知の物体であるマンホールのデータセットを作成し、テストデー

表 7 通常の物体検出モデルとラベルをオブジェクトに統一して学習したモデルによる、PascalVOC データセットに存在しないクラスのみ注目した COCO データセットのテストデータを用いた、未知の物体の検出性能の比較

検出モデル	$R_{0.25}$	$R_{0.5}$	$R_{0.75}$
通常 CenterNet	0.125	0.090	0.049
提案 CenterNet	0.146	0.104	0.058

タ 263 枚に対する、検出の性能を比較したものが表 8 である。ラベルをオブジェクトに統一したモデルは、通常の物体検出モデルと比較して、全ての閾値に対して、優れることはなかったが、評価値の Recall が高い傾向にある。

表 8 通常の物体検出モデルとラベルをオブジェクトに統一して学習したモデルによる、マンホールという未知の物体の検出の性能比較

検出モデル	$R_{0.25}$	$R_{0.5}$	$R_{0.75}$
通常 CenterNet	0.007	0.004	0.001
提案 CenterNet	0.009	0.005	0.000

また、通常の物体検出モデルによる未知の物体の検出性能が提案手法に競合する要因の一つに、CenterNet の性能が良くないことがある。ラベル通りに学習したモデルがマンホールを検出しているのは、偽陽性な出力であり、本来は誤りである。性能が良いモデルほど、偽陽性も減るため、性能が良くないモデルにおいては、未知の物体の検出が実現できる可能性を示す。

5.3 未知の物体の苦手なラベル分野

本研究の提案手法の物体検出モデルは、PascalVOC データセットの 20 クラスの画像データから物体らしさを抽出していると言える。しかし、PascalVOC データセットも多様ではなく、実世界の様々な物体と比較すると偏りがあると言える。図 6 は、4.3 での実験に用いたテストデータの物体らしさの検出結果である。図 6 では、人に対するオブジェクトの確信度は 0.70, 0.66, 0.48 と高いが、その他の物体に関してはリュックサックの確信度の 0.46 以外では、確信度が 0.3 を上回ることなく低く出力されている。これは PascalVOC データセットの学習データに大量の人のデータが存在し、その他の物体に関しては、PascalVOC データセットに存在しないクラスであることが原因だと考えられる。5.1 でのマンホールのデータセットに対する、検出結果もかなり低い。また、他のテストデータにも同様のことが言え、未知の物体に対する確信度は低くなる傾向がある。本研究は、未知の物体の検出を目的としているため、検出する上で、物体として認識する閾値は低く設定する必要がある。しかし、低く設定することで偽陽性も増えるため、適用させたい要件に適した物体として認識する閾値を各自設定す

る必要がある。

本研究では、PascalVOC データセットによって物体らしさを学習したが、COCO データセットのような物体の種類が多様で膨大なデータセットを活用することで物体らしさを検出するモデルの性能の向上が期待できる。

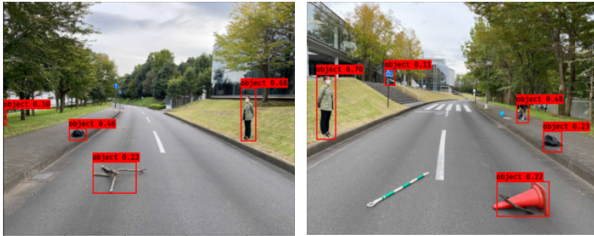


図 6 4.3 の実験に使用したテストデータによる、物体らしさの検出結果の一例である。

6. 展望

本研究の展望は、主に 3 つ存在する。1 つ目は、モデルの簡素化である。本研究の提案手法では、複数のモデルによって検出結果を得ており、処理の量や煩雑さが問題視できる。この問題は、古い物体検出モデルと似た問題であり、ワンショットの物体検出モデルが出現した時と同様の発展が期待できる。古い物体検出モデルは、領域を決める構造と分類する構造の二つに分けて検出を行っていたが、SSD[13]のような全て一つのニューラルネットワークモデルで推論を行えるワンショットの物体検出モデルが登場し、モデルの高速化や簡素化を実現した。そのため、本研究では独立して処理を行っていた、セグメント部分と物体検出部分、既知かどうかの分類を同時に処理できる構造を作り出すことで、上記の問題が改善することを期待できる。計算処理の量が改善されると、高速化や計算資源の限られたデバイスでも問題なく実行できるため、リアルタイム処理や IoT デバイスでの処理が可能となり、活用の幅が広がる。

2 つ目は、関連研究で紹介したような敵対的学習などを用いることで物体らしさを検出する性能を向上させることである。車道のセグメンテーションは高い精度で行えており、既知の物体に関しても学習済みの Yolov5 を用いているため高い精度で取り除くことができている。提案手法の性能面において、最もボトルネックになっているのは未知の物体を検出するモデル部分である。未知の物体の精度を向上させることでシステム全体の精度を向上が期待できるため、本研究の応用や発展させていく上で、最も注目すべき点だと言える。

3 つ目は、領域を制限しているため、処理コストを減少することが期待できる。本研究の提案手法では、画像全体に対して検出を行い、セグメンテーション結果による領域によって検出対象が取り除かれているが、事前にセグメンテーション処理を行い、その領域上での物体検出を考える

と、検出するべき画像領域は元画像よりもサイズが小さくなる。入力画像によって、画像の形状やサイズは異なるので、シンプルな物体検出モデルでは対処できないが、様々な形状の画像を処理できる物体検出モデルを実現させることができれば、入力のサイズが小さくなることから処理コストを大きく削減できる。この展望は、本研究の目的以外にも領域を限定した物体検出を求められる課題においては、実現すると利点が非常に大きい。

7. 結論

本研究では、日常的に起こりうる車道上の異状を問題視し、未知の物体の自動通知システムを目指す上での、車道上に限定した未知の物体の検出するための手法を提案した。本研究の目的である、未知の物体を検出する性能は、 $mAP_{0.5}$ では 0.720 の評価値であり、提案手法の有効性を示した。提案手法は、作成が困難である領域上にある未知の物体のデータセットを必要としないため、容易に実装が可能である。また、車道上に限らず様々な課題へと応用が可能である。今後の展望では、モデルの簡素化、未知の物体の検出精度の向上、推論時の計算コストの削減などがあり、取り組んでいく必要がある。

謝辞

本研究の一部は東京都江戸川区役所にご協力及び支援頂いた。

参考文献

- [1] Maeda, Hiroya and Sekimoto, Yoshihide and Seto, Toshikazu and Kashiyama, Takehiro and Omata, Hiroshi: *Road damage detection and classification using deep neural networks with smartphone images*, Computer - Aided Civil and Infrastructure Engineering 33.12 (2018): 1127-1141.
- [2] Alfarrarjeh, Abdullah and Trivedi, Dweep and Kim, Seon Ho and Shahabi, Cyrus: *A deep learning approach for road damage detection from smartphone images*, 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018.
- [3] Withanawasam, Jayani and Javanmardi, Ehsan and Kamijo, Shunsuke: *Top-down, Spatio-Temporal Attentional Guidance for On-road Object Detection*, 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2020.
- [4] Hendrycks, Dan and Gimpel, Kevin: *A baseline for detecting misclassified and out-of-distribution examples in neural networks*, arXiv preprint arXiv:1610.02136 (2016).
- [5] Jaiswal, Ayush and Wu, Yue and Natarajan, Pradeep and Natarajan, Premkumar: *Class-agnostic object detection*, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021.
- [6] Saito, Kuniaki and Yamamoto, Shohei and Ushiku, Yoshitaka and Harada, Tatsuya: *Open set domain adaptation by backpropagation*, Proceedings of the European Conference on Computer Vision (ECCV). 2018.

- [7] Wong, Kelvin and Wang, Shenlong and Ren, Mengye and Liang, Ming and Urtasun, Raquel: *Identifying unknown instances for autonomous driving*, Conference on Robot Learning. PMLR, 2020.
- [8] Zhou, Xingyi and Wang, Dequan and Krähenbühl, Philipp: *Objects as points*, arXiv preprint arXiv:1904.07850 (2019).
- [9] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian: *Deep residual learning for image recognition*, Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [10] Zhou, Zongwei and Siddiquee, Md Mahfuzur Rahman and Tajbakhsh, Nima and Liang, Jianming: *Unet++: A nested u-net architecture for medical image segmentation*, Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, 2018. 3-11.
- [11] Tan, Mingxing and Le, Quoc: *Efficientnet: Rethinking model scaling for convolutional neural networks*, International Conference on Machine Learning. PMLR, 2019.
- [12] Glenn Jocher, Alex Stoken, Ayush Chaurasia, JirkaBorovec, NanoCode012, TaoXie, Yonghye Kwon, KalenMichael, Liu Changyu, Jiacong Fang, Abhiram V, Laughing, tkianai, yxNONG, Piotr Skalski, AdamHogan, Jebastin Nadar, imyhxy, Lorenzo Mammana, AlexWang1900, Cristi Fati, Diego Montes, Jan Hajek, Laurentiu Diaconu, Mai Thanh Minh, Marc, albinxavi, fatih, oleg, and wanghaoyang0106: *ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support*, Zenodo, 2021.
- [13] Liu, Wei and Anguelov, Dragomir and Erhan, Dumitru and Szegedy, Christian and Reed, Scott and Fu, Cheng-Yang and Berg, Alexander C: *Ssd: Single shot multi-box detector*, European conference on computer vision. Springer, Cham, 2016.