

点群を用いた物体検出への説明可能なAIの導入*

Introduction of Explainable AI into Object Detection Using Point Clouds

黒木 理宏
Michihiro KUROKI

Explainable AI (XAI) is widely researched to enhance the interpretability of AI systems with black-box nature. XAI has made significant contributions in the field of object recognition, especially in image-based classification tasks. However, research on interpretability is relatively scarce when it comes to 3D object detection using point clouds from LiDAR sensors, which are commonly employed in autonomous driving applications. Given that these technologies are utilized in the safety-critical field, we think we should engage in the research of interpretability in 3D object detection using point clouds. In this paper, we propose a novel XAI method for 3D object detection and experimentally demonstrate its performance.

Key words :

Explainable AI, Object Detection, Point Clouds

1. はじめに

自動運転では安全な経路計画を行うため、周辺環境を高い精度で空間的に認識することが求められ、LiDAR センサによる高解像度な点群情報を用いた物体検出が注目されている。さらに近年ではAIの発展により、高度な点群の特徴量抽出を実現し、高い精度で高速に検出処理をする技術も研究されている。しかし、最新の高精度な検出技術であっても、誤った検出が存在する以上、AIを実際の車載環境に導入するのは難しい。なぜなら、AIはモデルの中身がブラックボックスであるため、出力結果の判断根拠に対する説明が難しいからである。なぜそのような検出が得られたかが説明できないのであれば、開発者がモデルの改善が難し

いだけでなく、搭載したシステムの顧客やユーザーの信頼まで低下させかねない。そこで近年では説明可能なAI (Explainable AI, XAI) として、AIの出力結果に説明性を付与して人間が理解しやすくする技術に注目が集まっている。

物体認識における説明可能なAIでは、モデルが出力した認識結果の判断根拠を、Fig. 1のように注目領域をハイライトする saliency map として提示する手法^{1) 4) 5) 7)}がよく取られる。現存する技術ではシステムの品質確保までは難しいものの、モデルがどのような認識を苦手とするかを開発者が理解し、開発へフィードバックすることができる。しかし、このような技術は画像上の物体のクラス分類のタスクに対するものが大部分を占めており、自動運転で多く活用される点群を用いた物

体検出のタスクに対してはほとんど研究されていない。そこで本稿では、従来の画像に対する説明可能なAIから、点群を用いた物体検出の説明可能なAIへの拡張について詳しく説明する。本稿では2章で関連技術の紹介を行った後、3章では点群を用いた物体検出に対する説明可能なAIについて述べる。4章ではそれらを用いた検証結果を示し、5章ではまとめについて述べる。



Fig. 1 Example of explainable AI in object recognition

2. 関連技術

2.1 点群を用いた物体検出

自動運転では周囲に存在する物体を直方体などの形状に当てはめて検出することが行われ、LiDARから得られる点群は物体の形状を3次元空間的にとらえることができるため、自動運転での物体検出に適していると考えられる。近年ではAIを用いて点群の特徴量を抽出する手法が用いられており、3次元上の格子空間に区切って格子毎に点群の特徴量を抽出する手法³⁾や、1点毎に特徴量を抽出する手法⁸⁾などが存在する。他にも検出の精度を高めるためにカメラから得られる画像情報を組み合わせたマルチセンサ物体検出器も存在する。画像と点群それぞれで得られた特徴量を組み合わ

せる手法⁹⁾もあれば、画像情報から物体の存在する領域を絞り、点群で詳細な形状を回帰する手法⁶⁾も存在する。本稿の4章の実験で用いたF-PointNets⁶⁾は初めに画像に対して物体検出を行い、得られた2次元物体検出領域を中心に3次元上の物体候補領域を生成する。物体候補領域内の点群のみを3次元物体位置の回帰に用いることで、領域を絞って効率的な点群処理を実現した手法となっている。

2.2 物体認識における説明可能なAI

主に画像上の物体のクラス分類に対する説明可能なAIの研究が進んでおり、saliency mapを説明結果として出力する手法が多く用いられる。そこでは入力画像の各画素において、予測結果への貢献度をヒートマップのように示している。貢献度を演算する手法はいくつも存在するが、Fig. 2のように大きく逆伝播ベース、活性化マップベース、摂動ベースのような手法に分けることもできる。逆伝播ベースの手法¹⁾では、入力画像から得られる予測スコアの情報をAIモデルの出力層から入力層まで逆方向に分解しながら伝播する計算を考える。その計算過程で得られるニューラルネットワークの勾配などの情報から、各画素が持つ予測結果への貢献度を得ることができる。活性化マップベースの手法⁷⁾では、畳み込みニューラルネットワークの最後の層が持つ特徴量マップが入力画像の特徴を集約していると仮定して説明結果に用いる。摂動ベースの手法⁴⁾では、先述した手法と異なりAIモデルの内部情報を用いない。代わりにマスク画像のような摂動を加えた入力サンプリングを用いて、モデルの出力や振る舞いの変化をもとに説明結果を出力する。

摂動ベースの手法はモデル内部の情報を用いないこ

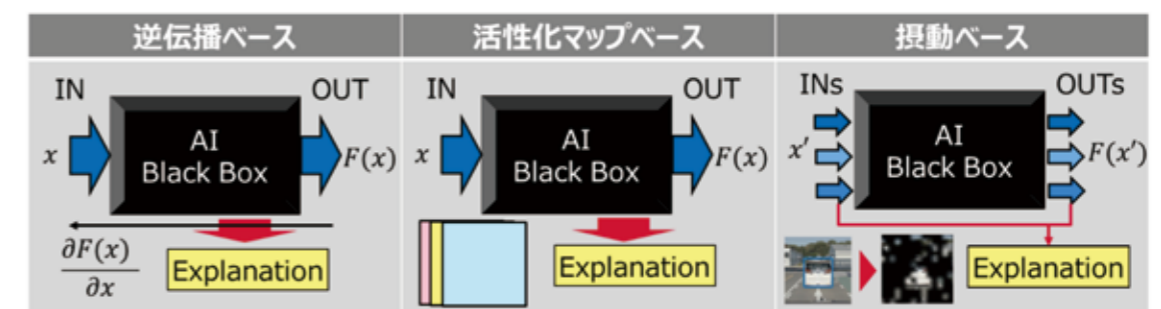


Fig. 2 Example of several types of explainable AI in object recognition

とから、サンプリングを必要とする計算面での課題はあるものの、様々なモデルに適用できる汎用性を持つ。そのため、画像上の物体のクラス分類向けのAIモデルに使えるだけでなく、位置特定も必要とする物体検出向けの複雑なAIモデルにも拡張が可能である⁵⁾。この拡張性に注目し、本稿では摂動ベースの説明可能なAIを自動運転における点群物体検出にまで拡張した手法について紹介する。

3. 手法紹介

研究例²⁾では、2D物体検出器向けの摂動ベース手法であるD-RISE⁵⁾を、点群を用いた3D物体検出器に拡張している。ここでは、物体検出器の入力点群に対してマスクなどの摂動を加え、得られる検出結果のスコアの変化をもとに各点群の重要度を求め、説明結果を出力することを考える。手法の全体像をFig. 3に示している。まず入力点群を X とし、AI物体検出器のモデルを f とすると、物体検出結果は $d \in f(X)$ と書くことができる。ここで d は、検出の領域やクラスなどを含んだベクトルとする。さらに、入力点群に摂動を与えるためのバイナリマスクを M とすれば、マスクされた入力点群は要素積 \odot を用いて $X \odot M$ として表すことができる。そして、判断根拠を表示するための説明対象の検出結果を d_t とすると、 d_t を検出するための各点群の重要度は、その点群が観測できている場合にどれだけ d_t に類似した検出ができるかの期待値で表すことができると考えられる。ここで、 d_t に類似した検出であ

るかの定量指標を検出類似度 S で表し、下記の式で定義する。

$$S \triangleq \max_{d_j \in f(X \odot M)} \text{sim}(d_t, d_j) \quad (1)$$

この時、 $\text{sim}(\cdot)$ は2つの検出結果の領域の重なり度を表すIntersection Over Union (IoU)や、物体クラスのコサイン類似度の積などを用いたものとする。詳しい式の導出は研究例²⁾を参照していただきたいが、各点群の重要度 I_f は、サンプリングされた入力点群に対するマスク M と、そこから得られた結果の検出類似度 S を用いることで、下記の式で表すことができる。

$$I_f = \frac{1}{Z} \sum_{i=1}^N S_i \cdot M_i \quad (2)$$

ここで Z は正規化のための係数、 N はサンプリングのマスク数とする。Fig. 3によると、車両から得られた点群を入力とした場合に、各点群の重要度を saliency map のように色付けした出力が得られていることが分かる。赤い点群は重要度が高い領域を示しており、この結果から検出において車両前方の部分重要視していることがわかる。

4. 検証結果

本章では、3章で紹介した説明可能なAIを実際に適用し、物体検出の判断根拠の結果について確認する。使用した物体検出器はF-PointNets⁶⁾とし、3次元上の物体候補領域内の点群を入力点群として扱っている。

また、学習と検証に使用したデータは弊社構内において弊社製のセンサーで取得したものとする。

データセットに存在した周辺車両の検出結果に対する判断根拠をFig. 4に示す。Fig. 4 (a)では簡単なシーンにおける2視点での結果を示しており、複数の車両で自車両に近い部分の面において強く反応しているということが伺える。LiDARの照射原理から考えて自車両に近い面に対象物体の点群が多く存在するという分布になっていると考えられ、AIでの学習においてこの特性を織り込んで物体検出を行っているということが推測できる。逆に反応が弱くなっている部分として共通しているのが、地面や周辺構造物の点群である。これは物体検出において、地面や周辺構造物が無関係な点群であるとAIモデルが正しく識別できていることの現れだと考えることができる。次にFig. 4 (b)の難しいシーンについて見てみると、上段では台車のようなものは車両とは無関係と識別できている様子が伺える。一方、中段では柱などに対して誤って注目し、下段では車両が複数並んでいるシーンにおいて誤って手前の車両に反応している様子も伺え、こういったシーンでの物体検出の難しさも理解できる。こういったシーンに対しては、学習データを増やして正しく特徴量を捉えられるようにするか、物体検出モデルの性能限界とみて他のモデルを試すかなどの柔軟な対応が必要となる。

以上のように、説明可能なAIから得られる物体検出の判断根拠を確認することによって、モデルが正しく特徴をとらえているか、苦手とするシーンはどこかなどを確認することができた。まだアプリケーションとしての確立は十分ではないが、追加学習用のデータ選定などのモデル改善にも活用できると考えている。

5. おわりに

本稿では物体認識における説明可能なAIを中心に、従来の画像での技術から点群を用いた物体検出向けの技術への拡張について説明した。データセットを用いて検証したところ、点群を用いた物体検出において重要とする領域を可視化することができ、物体検出が難しいシーンについても確認することができた。

点群を用いた物体検出に対する説明可能なAIはまだ登場したばかりであり、判断根拠の妥当性の確保や評価手法の確立など解決すべき課題も多い。実際のシステムでの物体検出の使われ方についても注意しながら、実課題へ適用できるような技術を実現していきたい。

参考文献

- 1) Sebastian Bach, et al. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. PLOS ONE, 10.7, (2015).
- 2) Michihiro Kuroki, et al. Visualization of Evidence by AI Object

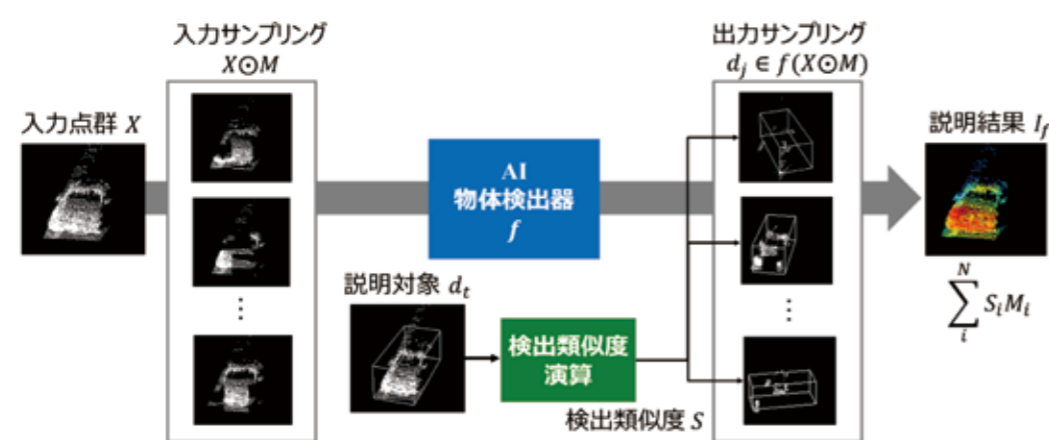


Fig. 3 Overview of explainable AI in 3D object detection using points clouds

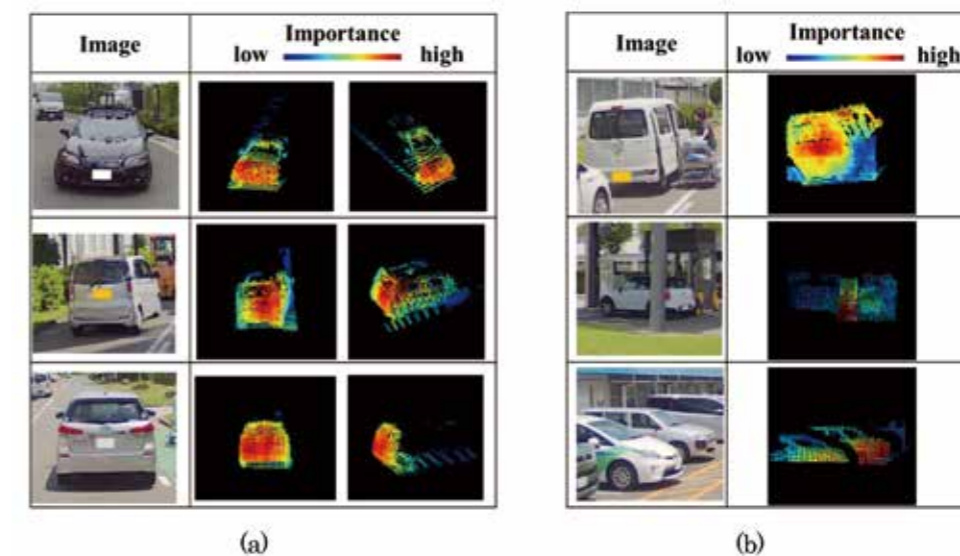


Fig. 4 Explanation results of explainable AI in 3D object detection
(a): easy example (b): hard example

- Detection Results Using 3D Point Clouds in Autonomous Driving. Transactions of Society of Automotive Engineers of Japan, 53.4, pages 802-807 (2022).
- 3) Alex H. Lang, et al. PointPillars: Fast Encoders for Object Detection from Point Clouds. In CVPR, pages 12697-12705 (2019).
 - 4) Vitali Petsiuk, et al. RISE: Randomized Input Sampling for Explanation of Black-box Models. In BMVC, page 151 (2018).
 - 5) Vitali Petsiuk, et al. Black-box explanation of object detectors via saliency maps. In CVPR, pages 11443-11452 (2021).
 - 6) Charles R. Qi, et al. Frustum pointnets for 3d object detection from rgb-d data. In CVPR, pages 918-927 (2018).
 - 7) Ramprasaath R. Selvaraju, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In ICCV, pages 618-626 (2017).
 - 8) Zetong Yang, et al. 3DSSD: Point-Based 3D Single Stage Object Detector. In CVPR, pages 11040-11048 (2020).
 - 9) Jin Hyeok Yoo, et al. 3D-CVF: Generating Joint Camera and LiDAR Features Using Cross-view Spatial Feature Fusion for 3D Object Detection. In ECCV, pages 720-736 (2020).

著者



黒木 理宏
くろき みちひろ

先進モビリティシステム事業開発部
修士 (工学)
自動運転における認識システム開発に従事