

実空間確率場表現に基づく三次元点群物体の推定手法の提案と評価

佐藤 広章[†] 荒川 伸一[†] 村田 正幸[†]

[†] 大阪大学 大学院情報科学研究科 〒565-0871 大阪府吹田市山田丘 1-5

E-mail: †{hi-sato,arakawa,murata}@ist.osaka-u.ac.jp

あらまし 遠隔地の実空間情報を利用した新しいネットワークサービスの登場が期待されている。実空間情報を利用したサービスの高度化のためには、実空間情報から実空間を理解することが重要であり、そのために機械学習を用いた物体識別技術の導入が進められている。本稿では、実空間の情報を機械学習等を用いて確定的に識別するのではなく、実空間を物体の確率的な重ね合わせの場として表現することに取り組んだ。多数のデータから実空間上に存在する物体の位置関係に対して隣接情報を取得し、実空間上の物体隣接にどのような相関関係があるかを統計的に求め、人間が経験的に備える実空間の知識情報を構築した。次に、機械学習による物体識別結果に対して、実空間の知識情報を加えて物体を推定する手法を考案した。評価の結果、機械学習による物体識別の確度が低い場合に推定精度が向上することがわかった。

キーワード 点群データ、確率場表現、物体識別、深層学習、セマンティックセグメンテーション、ベイズ推定

Proposal and evaluation of 3D-point object estimation method based on probability space representation

Hiroaki SATO[†], Shin'ichi ARAKAWA[†], and Masayuki MURATA[†]

[†] Graduate School of Information Science and Technology, Osaka University

Yamadaoka 1-5, Suita, Osaka, 565-0871 Japan

E-mail: †{hi-sato,arakawa,murata}@ist.osaka-u.ac.jp

Abstract New network services are expected to emerge using real spatial information in remote areas. For the advancement of services, it is important to understand real space from real spatial information, and for this purpose, object identification technology using machine learning is being introduced. In this paper, we have tried to represent real space as a field of probabilistic superposition of objects, instead of directly identifying real space information using machine learning techniques. We obtain neighboring information based on the positional relationships of objects in real space from a dataset used by the machine learning. Then, we propose a method of estimating objects by adding empirical knowledge information of real space to the object identification results. Our results show that predictions are changed by adapted empirical knowledge information and the estimation accuracy is improved when the confidence by the machine learning is not high.

Key words point cloud, random field representation, object detection, deep learning, semantic segmentation, Bayesian estimation

1. はじめに

遠隔地の実空間情報を利用した新しいネットワークサービスの登場が期待されている。例えば、自宅にいながら現地でのショッピング体験ができる遠隔ショッピングサービスなど、新しいネットワークサービスの登場が期待され、遠隔地で得られる情報から遠隔地の実空間理解や情報処理が求められている。具体的には、遠隔地で取得した実空間情報をネットワー

ク経由で転送し、アプリケーションを通して実空間情報から得たサービスをユーザに提供することが考えられる。これらのネットワークサービスを実現するには、実空間情報から実空間を理解することが重要である。例えば、車や通行人、建物などが混在している道路上のデータに対する物体識別、さまざまな種類の家具が配置されている屋内のデータに対する物体識別など、異なる状況下での物体識別の研究が多く行われている [1], [2]。これらは、自動車の自動運転や自動走行型ロボットの自律

制御など、実空間上の情報を高い精度で認識しなければならない環境において、必要なタスクの1つである [3]。また、サービスのリアルタイム性向上やネットワーク通信量削減のためには、高い識別精度に加えて限られたデータ量での識別や低遅延の処理が必要になる。

現在は、実世界情報を動画などの2次元データとして取得し、ディープラーニング等の機械学習技術を適用することで実世界情報をユーザに提示する研究が進められている [4]。しかし、実世界情報は2次元データに縮約されるため、例えば物体の重なりなどの立体的な情報を扱うことはできない。そこで近年では、3次元データに対する機械学習を用いた物体識別の研究が盛んに行われている [5], [6]。3次元データとして、Lidar センサーにより取得できる3次元点群データや、2つのカメラを用いる深度カメラにより取得する深度データなどが用いられる。既存の物体識別手法では、大量の学習データからデータに含まれる特徴を自動的に学習することで実装を行う手法である。しかし、既存の物体識別手法では、どのようなルールにより物体識別結果を導いているかを人間が理解するのは困難であることが指摘されている。また、我々人間が備えている実空間の知識情報、例えば台所とソファが隣接する可能性は低いなど、を無視した物体識別がなされる点に課題がある。

本稿では、実空間の情報を機械学習等を用いて確率的に識別するのではなく、実空間を確率的な場として表現することに取り組んだ。確率的な場とは、実空間にある物体を一意に識別するのではなく、物体のカテゴリの確率的な重ね合わせによって表現される場であり、例えば機械学習における識別精度向上に活用したり、遠隔地の実空間情報が欠落した際のデータ補完に活用することができる。多数のデータから実空間上に存在する物体の位置関係に対して隣接情報を取得し、実空間上の物体隣接にどのような相関関係があるかを統計的に求め、それを実空間の事前知識として与える [7]。さらに、事前知識と、観測等によって得られるデータに対して機械学習を適用して物体識別結果を重ね合わせることで、観測したデータに対する確率的な場としての確度を高める。実空間の統計的情報の取得には、3次元点群データを使用し、実空間の点群データを確率的な場として表現する。

さらに、確率場表現の利用例として機械学習による識別の精度改善を提示する。これにより、遠隔地の実空間情報をネットワークを介して転送し利用するサービスにおいて、データ量が少ない点群データに対して実空間情報の認識向上を目指す。

2. 実空間情報の確率場表現と取得方法

2.1 実空間情報

実空間情報として、実空間に存在する物体の位置やカテゴリの情報に着目する。例えば、屋内データであれば家具などの物体、屋外データであれば車や人などの物体の位置やカテゴリの情報を取り扱うこととする。これらの情報を扱うことで、例えば、自動運転や自律制御に必要な障害物などの検出、遠隔ショッピングサービスなどの新しいネットワークサービスにてユーザの欲しい商品情報などの認識など、様々な応用が可能で

ある。

実空間情報として3次元点群データを使用する。点群データとは、空間上の点の情報をまとめたデータであり、点ごとの空間座標情報 (X, Y, Z) と色情報 (R, G, B) の6パラメータが基本の構成要素となる。3次元点群データは、カラー画像と距離画像を同時に取得できる RGB-D カメラや測定対象物にレーザー光線を照射してレーザーが返ってくるまでの時間と照射角度から対象物の座標値とカラーカメラで取得した色情報を取得するレーザースキャナーによって測量するデータであり、複雑な形状を捉えることができる。

2.2 確率場表現

確率的な場とは、実空間にある物体を一意に識別するのではなく、物体のカテゴリの確率的な重ね合わせによって表現される場である。本稿では、多数のデータから実空間上に存在する物体の位置関係に対して隣接情報を取得し、実空間上の物体隣接にどのような相関関係があるかを統計的に求め、それを実空間の事前知識とする。実空間上の物体隣接にどのような相関関係があるかを統計的に求めることで、人間が備えている相関関係を数値化することができたり、事前知識として他の手法に組み込むことができる。今回は、2.2.1 節と 2.2.2 節で示す物体の存在確率分布と隣接確率分布を取得することで、物体カテゴリ間の物体数や隣接に対して、数値的な情報を獲得することを試みる。その際、物体の切り出しや位置の取得、隣接の取得が必要になるため、2.3 節で検出方法を説明する。

2.2.1 物体の存在確率分布

実空間情報から物体存在の統計的情報を取得する。実空間上に存在する物体情報を用いて、各カテゴリの物体数と物体の総数から各カテゴリの存在確率を算出することで、空間内の存在状況を取得する。

2.2.2 物体の隣接確率分布

実空間情報から物体隣接の統計的情報を取得する。実空間上に存在する物体情報を用いて、物体ごとの重心点間距離から隣接している物体を探索する。注目物体に対してどのカテゴリの物体が隣接しているかの確率を算出し、隣接と判断された物体を集計することで空間内の隣接状況を取得する。隣接確率は、対象の物体カテゴリと隣接の物体カテゴリの隣接数を対象の物体カテゴリの隣接検出数で割ることで算出する。

2.3 確率場表現の物体検出方法

本節では、実空間情報から物体の位置やカテゴリの情報を取得する方法を述べる。物体検出の方法としては、例えばクラスタリングを含めた教師無し学習や、深層学習を含めた教師有り学習があるが、本稿では密度ベースクラスタリングを用いる。密度ベースクラスタリングは、ポイントが集中するエリアと疎なエリアを分離し、ポイントが集中するエリアを物体として検出するアルゴリズムである。各カテゴリのみの点で構成された点群に対して密度ベースクラスタリングを実行することで、各カテゴリの物体位置を検出する。

2.3.1 密度ベースクラスタリング

3次元データに対するオープンソースライブラリーである Open3D [8] から、DBSCAN クラスタリング [9] を使用する。



図 1: 屋内 3 次元点群データ

ある点から半径 r 以内の点に対して探索を行い、到達可能点が閾値 n を超えている場合にクラスターと判断し、閾値 n を超えない点はノイズと判断するアルゴリズムである。

2.3.2 近傍点探索

2.3.1 節と同様に、3 次元データに対するオープンソースライブラリーである Open3D [8] から、高速最近傍点探索ライブラリー FLANN [10] を用いた KDTree の構築を使用する。KDTree は多次元空間上にある点を分類する空間分割データ構造であり、FLANN は次元数やデータ数によって変化する探索を高速化するライブラリーである。

3. 3 次元点群データセットを用いた確率場表現の例

3.1 実空間データ

屋内空間の公開データセットを使用して、実空間の確率場表現を取得する。本稿では、屋内 3 次元点群データセットである ScanNet [11] を用いる。この屋内 3 次元点群データセットは RGB-D カメラによって測量され、オフィス、アパート、バスルームなどシーンタイプにラベル付けされた 1613 部屋分のシーンに対して、座標情報と色情報を持つデータの集合であり、一例を図 1 に示している。

全 1613 部屋分のシーンに対して 1513 シーンが学習用データであり、この 1513 部屋分のデータセットを用いて確率場表現を得る。これらのシーンは学習用データセットとして、点群の空間座標と色の情報に加えて、表面の再構成やセマンティックセグメンテーションのアノテーションがつけられている。

図 2 に、ScanNet データセットに含まれる 1513 シーンのシーンタイプの出現分布を示す。縦軸がシーンタイプ名、横軸が各シーンタイプのシーン数を示す。21 種類のシーンタイプから構築される ScanNet データセットには、規模の小さいタイプ（バスルーム、クローゼット、ユニットバスなど）から大きなタイプの（アパート、教室、図書館など）までのさまざまな空間が含まれている。

各シーンには、部屋に設置されている物体に対応した点群オブジェクトが複数含まれている。以降、点群オブジェクトを物体と呼ぶ。物体には、その物体を表す物体カテゴリー識別子（以降、物体カテゴリー）が手作業によりアノテーションされている。表 1 は、ScanNet データセットの物体に対して付与さ

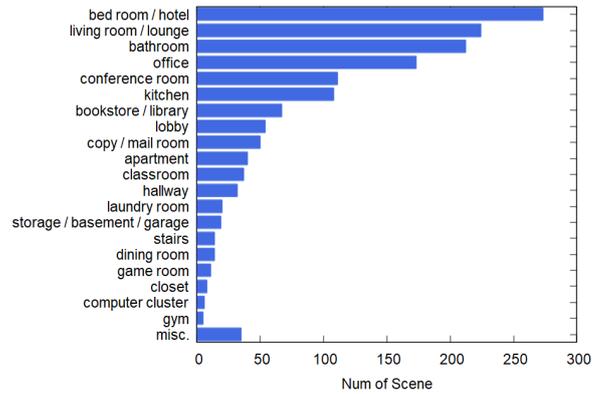


図 2: ScanNet データセットのシーンタイプの出現分布

表 1: 物体カテゴリー

Label	Category	Label	Category
0	wall	10	picture
1	floor	11	counter
2	cabinet	12	desk
3	bed	13	curtain
4	chair	14	refrigerator
5	sofa	15	shower curtain
6	table	16	toilet
7	door	17	sink
8	window	18	bathtub
9	bookshelf	19	other furniture

れた物体カテゴリーの分類である。other furniture には piano や display などが分類されている。なお、ScanNet データセットには壁 (wall) と床 (floor) の情報が含まれるが、壁や床の重心と物体の重心が離れることがあり隣接関係を適切に抽出することができないため、本稿では壁と床は除外して実空間情報を取得している。

3.2 確率場表現の例

2.3.1 節の閾値 n を 10、半径 r を 5 [cm]、物体隣接の重心点間距離を 1 [m] とし、2. 章で述べた方法により得た確率場の例を示す。

図 3 に確率場表現を取得した際の 1 つのシーンでの物体隣接関係を示す。点が物体の重心点、色が物体カテゴリー、直線が隣接を表している。この図から、2.3.1 節で示す方法で物体を取得する場合、点群データの取得時に生じるノイズや鮮明度の違いから、現実空間上では同一である物体が複数の物体に分割される（領域 a~c など）ことが読み取れる。また、隣接状況において、sink と counter と cabinet の隣接（領域 c）、desk と chair の隣接（領域 d）、ユニットバスでの cabinet と toilet と bathtub と shower curtain の隣接（領域 e）など、現実空間で見られる隣接状況が確認できるため、我々が生活を営む上での知識情報が取得できていることがわかる。

学習用データの全 1513 シーンに対する、物体存在確率分布と物体隣接確率分布の情報を表 2, 3 に示す。

表 2 から、table や chair、door などの一般的な部屋に置か

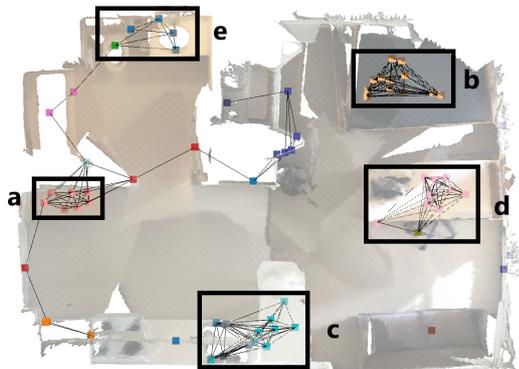


図 3: 物体隣接関係の例

表 2: 全シーンに対する存在確率 (% 表示)

Label	Category	Prob	Label	Category	Prob
0	wall	-	10	picture	2.35
1	floor	-	11	counter	2.11
2	cabinet	7.54	12	desk	6.26
3	bed	4.98	13	curtain	1.27
4	chair	19.61	14	refrigerator	1.28
5	sofa	3.44	15	shower curtain	0.41
6	table	22.52	16	toilet	0.81
7	door	8.56	17	sink	1.45
8	window	5.55	18	bathtub	0.57
9	bookshelf	3.55	19	other furniture	7.76

れている家具の存在確率が高いことが分かる。その次に存在確率が高いものとして、other furniture、cabinet、desk の家具が並ぶ。other furniture には piano や display などが分類されているため様々な家具が含まれたこと、cabinet や desk は図 2 で多く含まれているシーンタイプによくある家具であることが要因である。

表 3 から得た特徴を述べる。まず、表の対角の値が総じて大きいことがわかる。例えば、bookshelf (Label 9) は 92% であり、本棚の隣に本棚が置かれる性質が反映されている。また、shower curtain, toilet, sink, bathtub (Label 15-18) は低いパーセンテージとなっているが、機能上 1 つで十分なものであるにも関わらず、複数の物体が出現している。これは、2.3.1 節で示す方法で物体を取得する場合、点群データの取得時に起こったノイズや鮮明度の違いから、現実空間上では同一である物体が複数の物体に分割されるためである。また、ユニットバス関係 shower curtain, toilet, bathtub (Label 15, 16, 18) は相互的に隣接確率が高く、chair (Label 4) に着目すると、その隣接は、対角を除けば table (Label 6) 38% であり、続いて desk (Label 12) 8% である。加えて、toilet (Label 16) の隣に refrigerator (Label 14) は存在しない。これは、我々が理解している実空間上の特徴を反映していると考えられ、室内空間の特徴が確率的に表現できていると考えられる。なお、bed (Label 3) の隣に toilet (Label 16) が 1 件 (0.005%) 発生しているが、これは、シーンの中で壁を挟んだ物体の隣接が反映されているものであり、重心点距離の観点から隣接を取得する

方法から認識された隣接関係が抽出されたものであった。

シーンタイプ別の存在確率と隣接確率についても同様の傾向が得られている。加えて、例えばバスルームのシーンでは、shower curtain, toilet, sink, bathtub の物体数が多く、ユニットバスなどで見られる相互的な隣接である shower curtain と toilet, bathtub との隣接や、同様に水回り周辺での counter と sink の隣接などが多くあらわれており、隣接状況が正しく認識できていることが確認された。

4. 確率場表現に基づく事前知識を用いた物体推定手法

全 1513 シーンに対する確率場に対して、事前知識にもとづく物体推定手法への利用例を提示する。

4.1 既存手法

3 次元点群データに対する物体識別手法の既存手法として、SparseConvNet [13] モデルを使用する。SparseConvNet モデルは、深層学習を用いた物体識別のモデルであり、ScanNet データセットのセマンティックセグメンテーションタスクの精度ランキング [12] の上位に挙げられている。3 次元点群データに対するセマンティックセグメンテーションタスクは、各点に対して各物体カテゴリーのラベル付けを行うタスクである。物体カテゴリーそれぞれの確率が算出されるが、一般に確率が最大のもを物体の識別結果としてラベル付けがなされる。

Scannet データセットを用いた SparseConvNet の識別精度を表 4 に示す。セマンティックセグメンテーションの識別精度は IoU (Intersection over Union) と呼ばれる指標を用い、物体カテゴリーを正しく認識できた領域の割合を示す。既存手法では、80% を超える正解率で物体識別ができていた物体カテゴリーがいくつかあることに対して、20% 台の予測確率を持つ物体カテゴリーも存在する。

SparseConvNet などの深層学習にもとづく物体識別手法は、大量の学習データからデータに含まれる特徴を自動的に学習することで実装を行う手法であり、どのようなルールにより物体識別結果を導いているかを人間が理解するのは困難であることが指摘されている。また、我々人間が備えている実空間の知識情報、例えば台所とソファが隣接する可能性は低いなど、を無視した物体識別がなされる点に課題がある。

4.2 事前知識を考慮したセグメンテーション予測

取得した確率場表現を用いて、セマンティックセグメンテーション予測に対する物体推定を行う。セマンティックセグメンテーションの予測から、ベイズ推定 [14] により事前知識を含めた事後確率を算出する。

点群データ内のある物体 A に着目する。今、セマンティックセグメンテーションによる物体 A に対する物体カテゴリー $x \in X$ の予測確率を $p_A(x)$ とする。さらに、物体 A に対する k 個の隣接物体として $Z_{1:k} = Z_1, Z_2, \dots, Z_k$ の物体カテゴリーを $z_{1:k} = z_1, z_2, \dots, z_k$ とするとき、 $z_{1:k}$ を観察下での物体 A の事後確率 $P(x|z_{1:k})$ は、ベイズ推定を用いて、

表 3: 全シーンに対する隣接確率 (% 表示)

隣接 対象	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	71.9	0.1	3.4	0.3	1.2	1.4	0.8	0.4	0.5	7.7	3.2	0.3	1.6	0.0	0.8	3.9	0.1	2.7
3	0.1	95.3	1.0	0.0	0.2	0.2	0.4	0.2	0.3	0.0	1.3	0.4	0.0	0.0	0.0	0.0	0.0	0.7
4	1.1	0.3	48.1	0.2	38.9	0.5	0.9	0.2	0.1	0.0	8.0	0.3	0.1	0.0	0.0	0.0	0.0	1.5
5	0.7	0.1	1.3	80.5	11.4	0.2	0.6	0.0	0.4	1.4	0.5	0.3	0.0	0.0	0.0	0.0	0.0	2.5
6	0.2	0.0	17.7	0.7	79.8	0.2	0.3	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.7
7	2.1	0.2	2.1	0.1	1.6	81.9	1.1	1.1	0.6	0.5	1.1	0.6	0.5	0.6	0.8	0.7	0.3	4.1
8	1.5	0.7	5.0	0.4	4.2	1.4	75.2	0.5	0.5	0.9	4.0	1.7	0.1	0.0	0.1	0.5	0.0	3.3
9	0.9	0.3	1.6	0.0	0.1	1.6	0.6	92.6	0.2	0.0	0.7	0.1	0.0	0.0	0.0	0.0	0.0	1.2
10	3.1	1.6	1.7	0.9	4.2	2.6	1.8	0.6	74.9	0.5	3.2	0.3	0.2	0.6	0.4	0.5	0.0	2.8
11	27.1	0.0	0.2	2.0	0.3	1.1	1.6	0.0	0.3	50.0	0.0	0.2	3.4	0.0	0.1	10.8	0.0	2.9
12	3.1	1.2	23.2	0.2	1.2	0.7	2.1	0.3	0.5	0.0	64.6	0.3	0.1	0.0	0.0	0.0	0.0	2.6
13	3.8	4.3	9.5	1.6	6.6	4.7	11.8	0.7	0.6	0.6	4.1	45.4	0.5	0.0	0.1	0.5	0.0	5.0
14	11.7	0.0	1.2	0.1	0.3	2.6	0.3	0.0	0.3	7.3	0.7	0.3	67.8	0.0	0.0	2.7	0.0	4.7
15	1.9	0.0	0.0	0.0	1.3	17.7	0.2	0.0	4.2	0.0	0.0	0.0	0.0	23.0	19.0	2.7	25.0	5.0
16	15.1	0.1	0.1	0.0	0.6	10.7	0.7	0.0	1.3	0.4	0.0	0.2	0.0	8.3	25.1	8.4	11.0	18.1
17	30.4	0.0	0.3	0.0	1.3	3.4	2.1	0.0	0.7	24.3	0.0	0.3	2.8	0.5	3.3	22.0	0.8	7.8
18	2.4	0.0	0.0	0.0	1.1	6.9	0.4	0.0	0.0	0.0	0.0	0.0	0.0	17.6	17.8	3.1	45.9	4.7
19	3.8	1.0	6.3	1.4	6.1	3.8	2.5	0.7	0.6	1.2	3.7	0.6	0.9	0.2	1.3	1.4	0.2	64.1

表 4: SparseConvNet セマンティックセグメンテーション結果

Label	Category	IoU	Label	Category	IoU	
0	wall	0.7951	10	picture	0.2436	
1	floor	0.9509	11	counter	0.5920	
2	cabinet	0.6021	12	desk	0.5850	
3	bed	0.7874	13	curtain	0.6768	
4	chair	0.8907	14	refrigerator	0.4372	
5	sofa	0.8249	15	shower curtain	0.6142	
6	table	0.7151	16	toilet	0.9161	
7	door	0.5354	17	sink	0.6636	
8	window	0.5052	18	bathtub	0.8807	
9	bookshelf	0.7651	19	other furniture	0.4921	
					avg IoU	0.6737

$$P(x|z_{1:k}) = \frac{G(z_k|x) \times P(x|z_{1:k-1})}{\sum_{x' \in X} G(z_k|x') \times P(x'|z_{1:k-1})}$$

$$G(z_k|x) = p_{Z_k}(z_k) \times g(z_k|x) + (1 - p_{Z_k}(z_k)) \times (1 - g(z_k|x))$$

$$P(x|z_{1:0}) = p_A(x)$$

と求める。ここで事前知識として、その部屋のシーンタイプにおいて、カテゴリー a がカテゴリー b と隣接する確率を $g(b|a)$ とする。

4.3 評価結果

ベイズ推定を適用する前と後について性能指標を示す。

多クラス分類の性能指標に対して、Accuracy (正解率)、Precision (適合率)、Recall (再現率)、F1-score (F 尺度) を示す。Accuracy は、すべてのデータのうち正解だった割合、Precision はクラス L_i と予測したもののうちの正解の割合、Recall はクラス L_i データのうち正しく予測された割合、F1-score は Precision と Recall の調和平均である。

表 5: 手法の評価結果

(a) 各指標数値の比較

評価指標	Accuracy (正解率)	Precision (適合率)	Recall (再現率)	F1-score (F 尺度)
ベイズ推定 適用前	65.3%	60.1%	59.9%	60.0%
ベイズ推定 適用後	67.5%	70.8%	62.4%	66.3%

(b) 各ラベルの変化先

	適用後ラベル = 実ラベル	適用後ラベル ≠ 実ラベル
適用前ラベル = 実ラベル	6678 (88.7%)	850 (11.3%)
適用前ラベル ≠ 実ラベル	1099 (27.5%)	2896 (72.5%)

312 部屋の検証データの全ての物体に対して、セマンティックセグメンテーションに対する予測確率と事後確率の相関を正解ラベルデータを元に調べ、有効性を確認する。ベイズ推定を適用する前と後について性能指標を以下の表 5 に示す。Precision、Recall、F1-score は、各クラスに対する指標の平均を求めたものである。

表 5a から、Precision (適合率) は 10% 分ほど改善され、一定の効果を得ている。ただし Recall (再現率) は 3% ほど改善にとどまった。適合率が改善率が高いことは正しく補正された物体の数が多い傾向があると言えるが、再現率の改善率が低いことは、特定のクラス L_a の物体が別のクラスに補正された物体の数がある程度存在していると考えられる。適合率が大きく上昇したものの再現率が依然として低いクラスが多く、ベイズ推

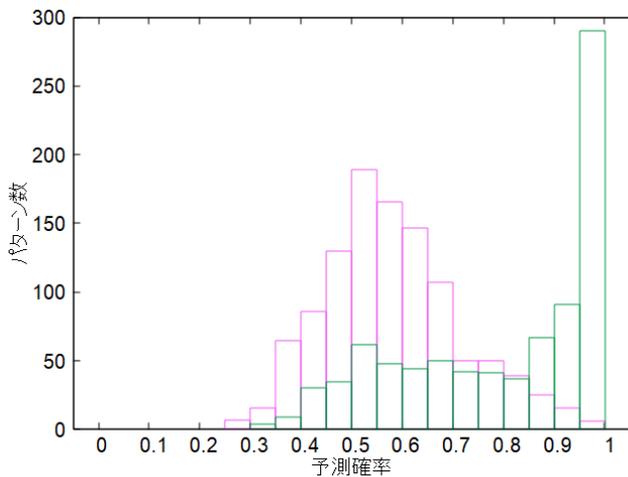


図 4: 予測ラベルが変化するパターンに対する予測確率分布

定を適用することで一定の物体予測の補正を行うことができたが、ベイズ推定では補正できなかった物体予測も多数存在していることが考えられる。

表 5b より、「事前ラベル \neq 実ラベル、事後ラベル \neq 実ラベル」のパターンが多い。図 4 に予測ラベルが変化したケースのヒストグラムを示す。予測ラベルの変化は、予測確率最大のラベルと正解ラベルが不一致の状態から正解ラベルを導いたパターンと、一致の状態から間違いラベルを導いたパターンがあり、図中ではそれぞれピンク色の領域と緑色の領域で示している。

図 4 より、正解ラベルを導いたパターンはベイズ推定前の予測確率が 0.5~0.6 のあたりに集中していることが分かる。また、間違いラベルを導いたパターンは、ベイズ推定前の予測確率が 0.9 以上に集中していることが分かる。そのため、予測確率が 0.9 以上である物体にはベイズ推定を適用せず、セグメンテーション予測の結果を保持することが考えられる。

5. まとめと今後の課題

本稿では、実空間の情報を確率的な場として表現することに取り組んだ。実空間の情報のデータとして、屋内 3 次元点群データセットである ScanNet から、実空間上に存在する物体の位置関係に対して隣接情報を取得し、実空間上の物体隣接にどのような相関関係があるかを統計的に求め、実空間の確率場を取得した。次に、取得した確率場を用いて、事前知識にもとづく物体推定手法への利用例を提示した。評価の結果、事前知識にもとづく物体推定手法を行うことで、Accuracy、Precision、Recall、F1-score それぞれで一定の推定精度向上が確認された。また、既存手法によるセグメンテーション予測が十分な確度を得られていない場合に、事前知識にもとづく物体推定が有効であることも確認された。

本稿の評価では、実空間情報の確率場表現を物体推定手法に組み込めることを確認したが、実空間上で利用する物体推定手法としては改善の余地があると考えられる。加えて、例えば、他の屋内データセットや屋外データセットなど、複数のデータセットから実空間情報の確率場を取得することで、より正確な

実空間情報を取得できていると考えている。また、セマンティックセグメンテーションを実行することで、その予測結果を事前知識としてアップデートをしていくことも考えられる。

文 献

- [1] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.
- [2] T. de Vries, I. Misra, C. Wang, and L. van der Maaten, "Does object recognition work for everyone?," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 52–59, June 2019.
- [3] L. Wang, X. Fan, J. Chen, J. Cheng, J. Tan, and X. Ma, "3D object detection based on sparse convolution neural network and feature fusion for autonomous driving in smart cities," *Sustainable Cities and Society*, vol. 54, p. 102002, Mar. 2020.
- [4] S. Takagi, J. Kaneda, S. Arakawa, and M. Murata, "An improvement of service qualities by edge computing in network-oriented mixed reality application," in *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, pp. 773–778, Apr. 2019.
- [5] J. Zhang, X. Zhao, Z. Chen, and Z. Lu, "A review of deep learning-based semantic segmentation for point cloud," *IEEE Access*, vol. 7, pp. 179118–179133, Dec. 2019.
- [6] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *International Journal of Computer Vision*, vol. 128, pp. 261–318, Oct. 2019.
- [7] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 282–289, June 2001.
- [8] "Open3D - A Modern Library for 3D Data Processing." <http://www.open3d.org/>. (Accessed: 2020-11-18).
- [9] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 226–231, Aug. 1996.
- [10] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pp. 331–340, Feb. 2009.
- [11] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3D reconstructions of indoor scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2432–2443, July 2017.
- [12] "ScanNet Benchmark." http://kaldir.vc.in.tum.de/scannet_benchmark/. (Accessed: 2020-10-05).
- [13] B. Graham, M. Engelcke, and L. van der Maaten, "3D semantic segmentation with submanifold sparse convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9224–9232, June 2018.
- [14] B. Ristic, C. Gilliam, M. Byrne, and A. Benavoli, "A tutorial on uncertainty modeling for machine reasoning," *Information Fusion*, vol. 55, pp. 30–44, Mar. 2020.