基盤モデルによる単眼画像深度推定と画像領域分割を用いた屋外自律移動ロボットの誘導

○佐藤 康晴（九州大学），松本 耕平（九州大学），倉爪 亮（九州大学）

Navigation for Outdoor Autonomous Mobile Robot Using Monocular Image Depth Estimation and Image Segmentation Based on Foundation Models

○ Kosei SATO (Kyushu University), Kohei MATSUMOTO (Kyushu University),

and Ryo KURAZUME (Kyushu University)

Abstract: This study proposes an autonomous navigation system for outdoor robots that relies solely on camera-based sensing, using a minimal sensor configuration without range sensors. By applying one-shot semi-supervised learning in real-world environments, the system segments images into traversable and non-traversable regions. In addition, depth estimated from monocular images enables inference of traversable areas in 3D space. Point clouds generated from these monocular depth estimates are further utilized to perform obstacle avoidance. This paper presents the results of applying the proposed techniques to a mowing robot, including recognition of trimming areas and obstacle detection during weed-trimming tasks.

1. 緒言

近年，高齢化や過疎化の進行による人手不足の解決策として，様々な分野へのロボットの導入が期待されている．特に，屋外移動ロボットは，農作業や土工作業など人手では過酷であったり危険が伴うような作業において代替を担うことが期待されている．

屋外移動ロボットに作業を代替させる際に重要な点として，ロボットの正確な位置推定，走行領域の指定，経路生成および経路追従，障害物の回避のための周辺環境の把握が挙げられる．我々はこの中でも走行領域の指定について注目し，本稿では，走行領域を適切に指定する必要がある具体的な作業として，農地などでの草刈り作業を想定する．

草刈り作業を自律移動ロボットによって実現する先行研究において，走行領域を指定する方法として，航空写真で範囲を指定する手法がある[1)2)3)]．しかし，この手法では航空写真が古く，実環境と一致しない場合があり，実環境を反映して走行領域を指定する場合には正しく指定することが難しい．加えて，航空写真は広域写真であることから，細かな範囲を指定することも難しい．

本研究では画像処理の基盤モデルを活用し、一枚の画像から走行領域として指定したい領域を選択して半教師あり学習を行う。ロボットの移動によって得られる新たな視点からの画像に対しても推論を実施し、得られた画像ベースの範囲情報を、ロボットから取得したオドメトリおよび画像の深度推定による深度情報と統合することで、走行領域の 3 次元点群を生成する。これにより、実環境を反映した細かな範囲指定を 1 つのラベル付きデータで実現する手法を提案する。

本研究では具体的に以下の要求仕様を満たすシステムの実現を目指す．

- 草刈り領域のような，現場によって対象が大きく変化するような一貫性が無い領域を推定するために，現場での学習が可能であること
- 現場での学習の際に少量データからでも学習できること
- 現場での学習の際にラベル付け作業の労力ができるだけかからないこと
- 現場での学習が現実的な時間で可能であること

また，具体的な問題設定として，背の高い草と背の低い草のある環境にて，背の高い草のある範囲を走行領域として判別する状況を考える．

2. システム構成

本項では，走行領域の 3 次元点群を得るために開発したシステムの構成について説明する．

2.1 全体構成

本研究では，カメラから得られた画像から走行領域を示す 3 次元点群を生成するまでのシステムを ROS 2 Humble 上にて構成した．システム全体の構成について **Fig. 1** に示す．

システム起動後，カメラから受け取った 1 番目の画像に対し，GUI を通して手動で走行領域と非走行領域を指定することで正例と負例のラベル付きデータを作成し，画像領域分割についての半教師あり学習を行う．周辺の環境が十分カメラ上に映るようにロボットを移動することで得られていく画像ごとに推論を行い，走行領域と非走行領域に領域分割をする．走行領域と判定された部分の深度を推定して得られる走行領域の 3 次元点群を逐次追加していき，全体の走行領域として指定したい範囲の 3 次元点群を得る．

2.2 深度推定

本研究では，3 次元点群を作成するために画像の深度を取得する必要があるため，視覚基盤モデルの一種である，絶対深度推定用にファインチューニング済みの Depth Anything V2[4)] を，エッジデバイスでの動作を考慮して TensorRT により推論を高速化した状態で用いている．

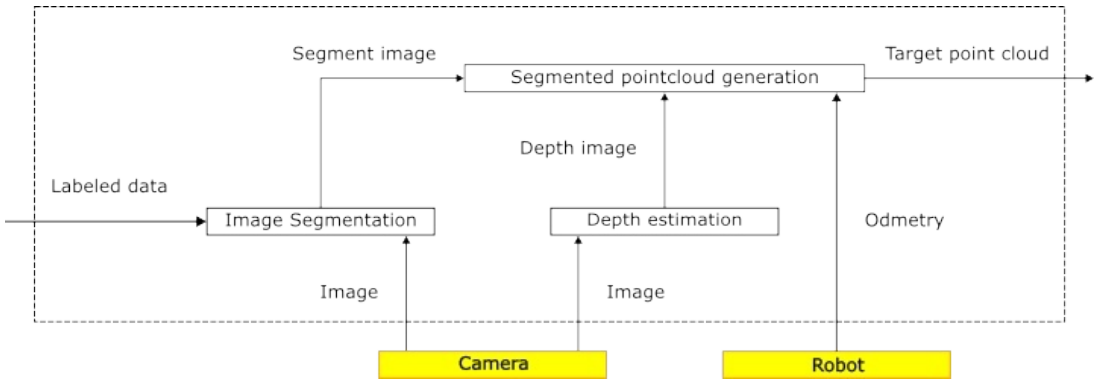


Fig. 1 A sequence of software components that generate 3D point cloud data of the target area. The black-bordered blocks enclosed by dotted lines indicate software elements, while the orange-bordered blocks indicate hardware elements.

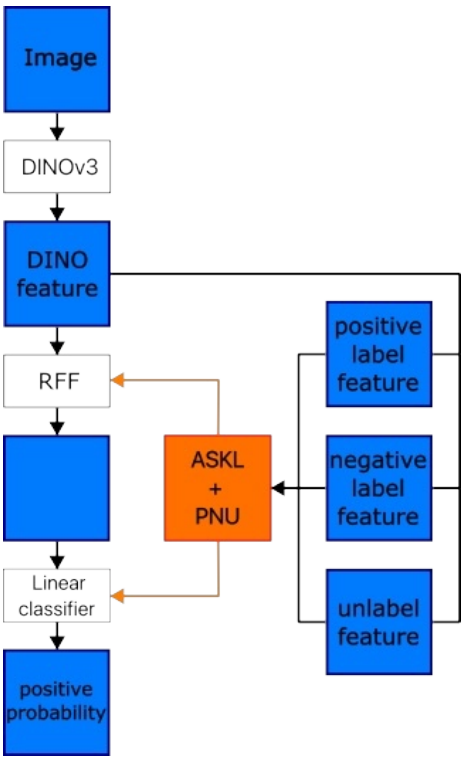


Fig. 2 System configuration of image segmentation

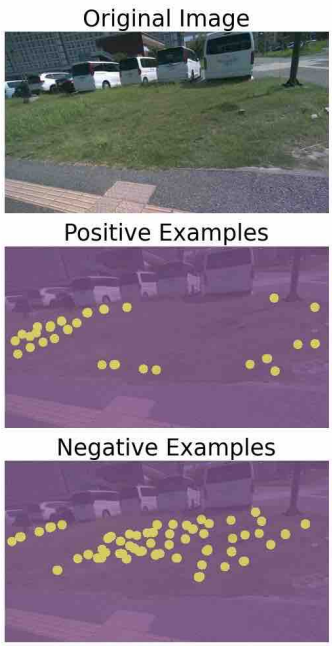


Fig. 3 Label data and original images for the comparison

画像の深度を得る方法には，主に

- RGBD カメラを利用する方法
- 画像と同期した 3D-LiDAR から取得する方法
- 画像からの深度推定による方法

が挙げられる．その中でも，本システムにて用いる際に要求される，屋外のロボットへの適用と細かい領域においても深度が必要であるため遠方でも高解像度の深度を得られること，ロボットから得られるオドメトリ情報を加えて実際の環境を反映した 3 次元点群を生成する必要があるため絶対深度が得られること，ロボットへの導入を考慮し軽量化されたモデルが必要であること，という条件を満たす基盤モデルによる深度推定より深度情報を得る方法を採用した．

2.3 画像領域分割

提案する画像領域分割手法は，DINOv3[5] と，ランダムフーリエ特徴 (RFF)[6] による特徴抽出，Positive-Negative-Unlabeled Learning (PNU 学習)[7] と Autmated Spectral Kernel Learning (ASKL)[8] を組み合わせて学習する線形識別器によって構成される．本研究で用いる画像領域分割の処理の構成を **Fig. 2** に示す．

2.3.1 DINOv3

DINOv3 は視覚基盤モデルの一種であり，任意の RGB 画像から，セマンティックセグメンテーションや単眼深度推定などの様々なタスクに応用可能な汎用性の高い特徴量を抽出できる．本研究では，少量データを用いてかつ，学習時間を抑える軽量な学習においても高い汎化性能を実現するために DINOv3 を用いる．

2.3.2 Positive-Negative-Unlabeled Learning

現場でのラベル付け作業の労力の低減のため，ラベルがデータの一部にしか付与されていない場合を許容するために PNU 学習を採用する．PNU 学習は，正例と負例のラベル付きデータおよびラベルなしデータを活用して半教師あり学習を行うための学習手法であり，ラベル付けが不完全なデータからの学習が可能である．本研究では，正例として走行したい領域（草刈りを行いたい領域などを想定），負例として走行してほしくない領域を少数指定して学習を行う．

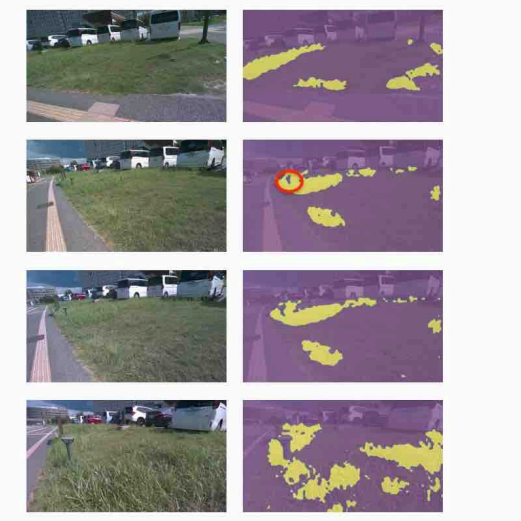


Fig. 4 Results of segmentation using ASKL + RFF + PNU Learning

Fig. 5 Results of segmentation using RFF + PNU Learning

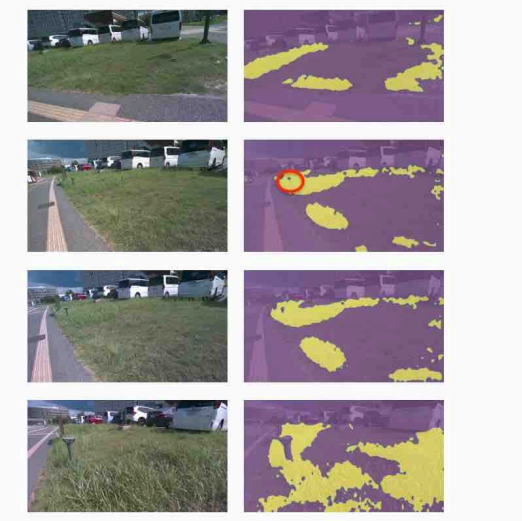


Fig. 6 Results of segmentation using PNU Learning

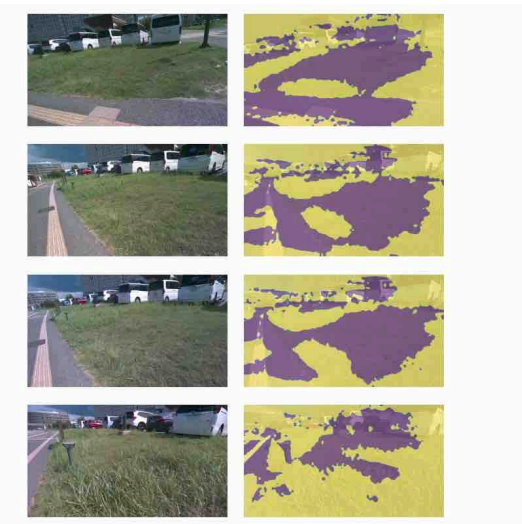


Fig. 7 Results of segmentation using Logistic Regression

2.3.3 ランダムフーリエ特徴

ランダムフーリエ特徴 (RFF) はカーネル法におけるカーネル関数をフーリエ変換とモンテカルロ近似を用いて近似することによって得られる関数によって計算される．近似によってカーネル法の恩恵を受けつつ，通常のカーネル方と比較して計算量を大幅に減らすことができる．本研究では RFF は領域分割性能の向上を目的として採用し，後述の Automated Spectral Kernel Learning と関連して，非定常スペクトルカーネルを用いる．

2.3.4 Automated Spectral Kernel Learning

ASKL は非定常スペクトルカーネルを採用することと，固定値ではなくではなく学習によりパラメータを最適化することで，入力データと出力データに対応する形でパラメータを設定することができる．従来の定常スペクトルカーネルを用いる場合は，パラメータ決定が入力データおよび出力データどちらにも依存しないことによって性能に限界があったが，ASKL によってこの課題点を克服できる．本研究では領域分割性能の向上のために ASKL を採用している．

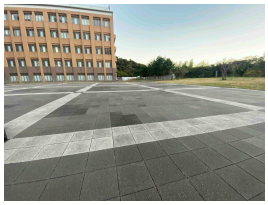
2.3.5 比較結果

本研究で提案する画像領域分割手法について各コンポーネントを除いた手法と参考としてロジステック回帰を比較対象とした．草刈りタスクを適用例として，背の高い草の精度を比較する．この際，学習時に部分ラベル付きデータとして与えた正例と負例のマスク画像と元画像を **Fig. 3** に示す．

- PNU Learning + RFF + ASKL
- PNU Learning + RFF
- PNU Learning
- Logistic Regression

各々の手法を一連の４フレームに適用し，予測確率を 50%を閾値にして領域分割した画像と元画像を並べた結果を **Fig. 4**，**Fig. 5**，**Fig. 6**，**Fig. 7** に示す．

Fig. 7 から，背の高い草以外の多くの領域を誤判定してしまうため，ロジスティック回帰では正しく領域分割できないことは明らかである．Fig. 5 では，背の高い草の領域であるのにも関わらず正しく判定できていない部分が多く確認された．したがって，DINOv3 と



(a) Test field

(b) Tripod for camera as an obstacle

Fig. 8 Settings for the obstacle avoidance experiment

(a) Front side

(b) Rear side

Fig. 9 Crawler robot

RFF による特徴抽出に基づく PNU 学習手法の有効性が低いことが明らかになった．DINOv3 による特徴抽出と PNU 学習のみの手法と，DINOv3 と RFF による特徴抽出に基づく PNU 学習および ASKL を組み合わせた手法により得られる結果を比べた際，PNU 学習のみの方が，全体的に背の高い草の領域を正しく判定できているが，Fig. 6 の赤丸に囲われた部分にある遠くの電灯の柄の部分を背の高い草の領域であると誤判定している．一方で，PNU 学習と ASKL を組み合わせた手法では，Fig. 6 の赤丸に囲われた部分にある遠くの電灯の柄の部分を背の高い草の領域ではないと正しく判定できている．本システムでは，走行領域と判定された部分を逐次追加して全体の走行領域の 3 次元点群を生成する流れであるため，1 枚の画像について走行したい領域を非走行領域と誤判定しても，別の角度からの画像からでの修正が可能である．したがって，走行してほしくない領域を走行領域として誤判定しないことを優先して PNU 学習と ASKL を組み合わせた手法を採用した．

3. 実験

本項では，走行領域の 3 次元点群を得るシステムの具体的な適用例として挙げた草刈りタスクについて，本システムを適用した実験結果について示す．

3.1 障害物回避実験

Depth Anything V2 より推定される深度が実環境で用いることにおいて有効であるか，実環境での障害物回避実験を通して確認した．

3.1.1 実験環境

九州大学ウエスト 2 号館前の **Fig. 8(a)** に示す，開けた広場において行った．障害物として **Fig. 8(b)** に示す三脚を用いている．

3.1.2 ロボットの構成

障害物回避実験で使用したロボットを **Fig. 9** に示す．ハードウェア構成は，クローラーロボットに制御するコンピューターとして Jetson AGX Orin を用い，カメラには RealSense D457 を用いている．ソフトウェアについては，ROS 2 Humble 上に構成し，姿勢情報と画像情報を RealSense D457 より取得し，画像情報から深度推定を行って得た周辺環境の 3 次元点群，姿勢情報およびロボットより得られる wheel odometry を ROS 2 のパッケージである Navigation 2 にそれぞれ渡し，ロボットのナビゲーションを行っている．移動目標地点については Rviz2 を通して指定している．ここでは RealSense D457 を用いているが，これより直接得られる深度情報は用いていない．

3.1.3 結果

実際に動作させた所，**Fig. 10(a)** に示すように，障害物を点群で捉えており，障害物として Navigation 2 も捉えられている．この時，Navigation 2 の経路生成で得た経路およびコストマップを Rviz 上に表示したものを **Fig. 10(b)** に示す．青点が移動の始点，赤点が移動の終点，橙色の点が障害物の位置，緑の線は生成した移動経路である．実際に動作した結果，ナビゲーションの終点と指定した地点に 1m ほどの差異が見られた．その原因としては，自己位置推定の情報源が wheel odometry と姿勢情報のみと少なかったためであり，障害物を認識し，経路自体は避けるように生成されていたため，今回採用した深度推定を用いても周辺環境の 3 次元での認識は問題にないと考えられる．

3.2 走行領域推定実験

本研究にて開発した走行領域の 3 次元点群を得るシステムを用いて，草刈りタスクを適用例とし，背の高い草の領域のみの 3 次元点群として生成できるかを確認した．

3.2.1 実験環境

九州大学ウエスト 1 号館横の **Fig. 11** に示す場所において行った．

3.2.2 ロボットの構成

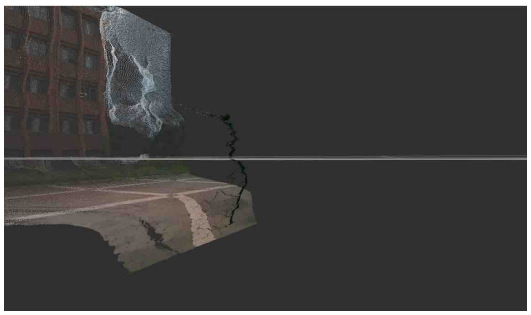
本実験では，ソフトウェアについては，ROS 2 Humble 上に構成し，画像情報を RealSense D457 より取得し，RTAB-Map へ渡し visual odometry を得る．それにより得られた画像情報および visual odometry を本研究にて開発したシステムに渡す構造になっている．ここでは RealSense D457 を用いているが，本カメラから直接得られる深度情報は用いていない．

3.2.3 結果

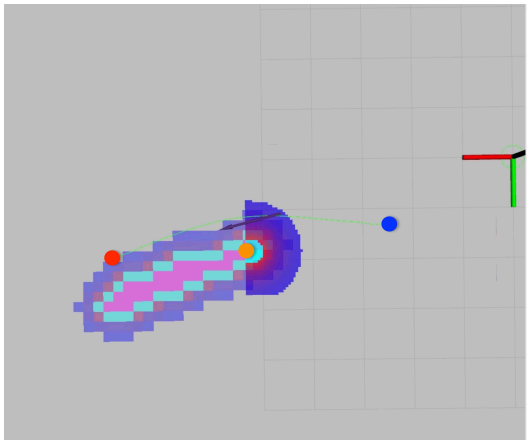
本研究で開発したシステムにより走行領域の 3 次元点群を得ていく時系列での流れを **Fig. 12** に示す．右は画像に対しての画像領域分割結果が，左はその時点までで推定された走行領域の 3 次元点群を示している．

4. 結言

本研究では，現場で作成した 1 セットの部分ラベル付きデータを元に走行領域と非走行領域について画像領域分割を行い，深度推定によって得られた深度と組み合わせることで，走行領域の 3 次元点群を得る一連のシステムを開発した．本システムは，現場によって対象が大きく変化するような一貫性が無い領域を推定するために，現場において，ラベル付け作業の労力が



(a) 3D point cloud of the surrounding environment



(b) Result of path planning

Fig. 10 Results of the obstacle avoidance experiment

少ない少量のラベル付を行い，かつ，学習が現実的な時間で可能であるという条件を満たすことを目標として開発された．草刈りタスクを適用例として，本システムを用い，背の高い草を走行領域とする 3 次元点群が生成されることを確認した．現在，画像領域分割として用いている手法では，人による 1 セットの部分ラベル付きデータのみで学習を行っている．そのため，ロボットの移動により，視点が大きく変化した場合，始めのラベル付きデータを作成する画像には写っていない物体が写ることで，高い精度で走行領域を判定できなくなる場合が考えられる．そのため今後は，画像領域分割に関わる学習を，走行領域推定を行いながら継続的に行うように変更を加える必要がある．具体的には，新たにカメラより画像を得るたびに，新たな画像の中で，本システムを用いて推定された走行領域に対応している場所を写している部分を正例としてラベル付けしたデータを用いて，追加学習を行う手法を検討していきたい．

Fig. 11 Test field for the area estimation experiment

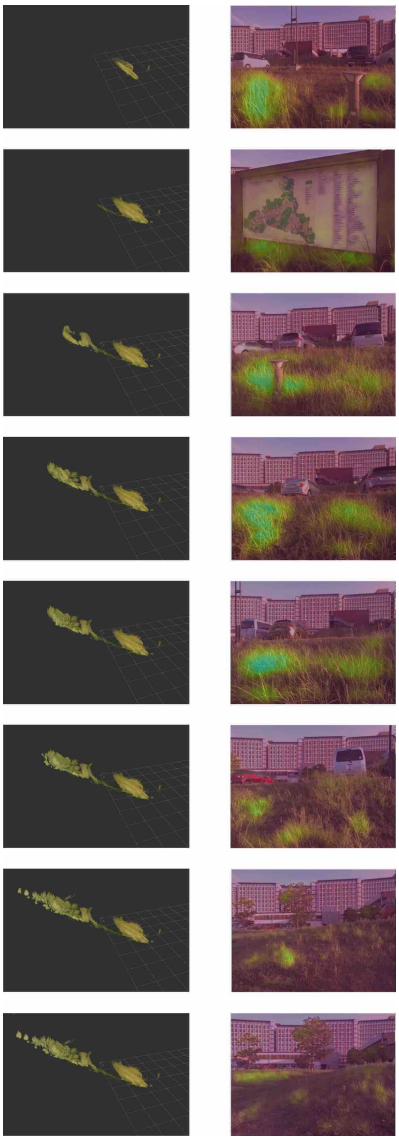


Fig. 12 3D point cloud of the target area (left side), and results of area estimatioon (right side)

参考文献

[1] 林 拓真, 大城 孝弘, 渡邉 崇, 下窪 竜, 小玉 尚人, 倉爪 亮: 高精度 GNSS を用いた自律移動草刈りロボットの開発, 日本機械学会ロボティクスメカトロニクス講演会 2021 (2021), 1P2–A12.

[2] 林 拓真, 大城 孝弘, 渡邉 崇, 下窪 竜, 小玉 尚人, 倉爪 亮: 高精度 GNSS を用いた自律移動草刈りロボットの開発, 第 22 回計測自動制御学会システムインテグレーション部門講演会 SI2021 (2021), 1G3–05.

[3] 松本 耕平, 大城 孝弘, 渡邉 崇, 下窪 竜, 小玉 尚人, 倉爪 亮: 高精度 GNSS を用いた自律移動草刈りロボットの開発 -第三報 QZSS と Visual SLAM カメラによる位置推定と経路追従実験-, 日本機械学会ロボティクスメカトロニクス講演会 2023 (2023), 2P1–B03.

[4] L. Yang, B. Kang, Z. Huang, Z. Zhao, X. Xu, J. Feng, and H. Zhao: Depth Anything V2, (2024), arXiv: 2406.09414.

[5] O. Siméoni, H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M.

Ramamonjisoa, F. Massa, D. Haziza, L. Wehrstedt, J. Wang, T. Darcet, T. Moutakanni, L. Sentana, C. Roberts, A. Vedaldi, J. Tolan, J. Brandt, C. Couprie, J. Mairal, H. Jégou, P. Labatut, and P. Bojanowski: DINOv3, (2025), arXiv: 2508.10104.

[6] A. Rahimi and B. Recht: Random features for large-scale kernel machines, Advances in International Conference on Neural Information Processing Systems (NeurIPS) (2007), pp. 1177–1184.

[7] T. Sakai, M. C. du Plessis, G. Niu, and M. Sugiyama: Semi-supervised classification based on classification from positive and unlabeled data, Proceedings of the International Conference on Machine Learning (ICML) (2017), pp. 2998–3006.

[8] J. Li, Y. Liu, and W. Wang: Automated Spectral Kernel Learning (2020), pp. 4618–4625.