セマンティックセグメンテーションを用いた深層強化学習による自律移動の獲得

○丸山祐矢　古川弘憲　村瀬卓也　山内悠嗣　山下隆義　藤吉弘亘 (中部大学)

1. はじめに

ロボットによる動作の自動獲得は，ルールベースのアプローチではなく，深層学習を用いた学習ベースのアプローチが利用されている．文献 [1] は，自動運転を対象とし，画像を入力，ハンドルの操舵角を教師信号として，畳み込みニューラルネットワークによる End-to-end 学習を実現し，適切な動作を推定する．深層強化学習を用いた文献 [2] では，14 台のロボットマニピュレータで 80 万回の把持動作を収集し，強化学習により把持動作を自動獲得している．前者の手法は教師あり学習であり，教師信号として実空間における運転動作と画像を大量に収集する必要がある．一方，強化学習である後者の手法は，教師信号を必要としないが，ロボットを実空間で動作させる必要があり，学習に多大な時間を要する．特に，強化学習を用いて自律移動のための動作を実機で獲得するには，衝突等の学習に必要となるイベントが伴うため，その学習コストは高い．この問題を解決する方法として，シミュレータの利用が考えられる．OpenAI Gym[3] は，強化学習のためのシミュレーションプラットフォームであり，深層強化学習に利用され始めている．このような仮想環境における学習はコストが掛からず，シミュレータを用いることで実時間以上の速度で動かすことによる高速な強化学習が可能である．しかし，シミュレータ環境における入力とシーンを撮影した実空間の入力データの間には大きな隔たりが存在する．そのため，シミュレータ環境で深層強化学習により学習したネットワークモデルを用いて実機で評価すると上手く動かないという問題が発生することがある．

そこで，本研究ではこれらの問題を解決するため，セマンティックセグメンテーションを用いた強化学習による自律移動の自動獲得を実現する．実画像で学習したセマンティックセグメンテーションの結果を入力として強化学習することで，シミュレータ・現実空間の間での入力データの隔たりを無くすことが期待できる．これにより，高速かつ低コストなシミュレータ環境での学習結果を用いて，実世界でのロボット制御を実現することが可能となる．従来の教師あり学習と強化学習と提案するアプローチの学習の流れを図 1 に示す．

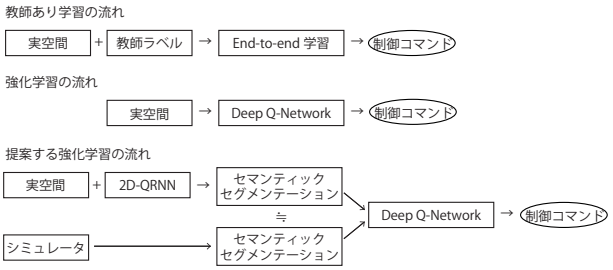


図 1: 教師あり学習と強化学習と提案するアプローチの学習の流れ

2. 2D-QRNN

本章では，セマンティックセグメンテーションで用いる 2D-QRNN について述べる．セマンティックセグメンテーションは，画像に含まれる物体をピクセル単位で識別する問題である．深層学習である畳み込みニューラネットワーク (CNN) を用いた手法には，Fully Convolutional Networks[4] や SegNet[5]，DAG-RNNs[6]，CNN に 2D-QRNN を導入した手法 [7] などがある．DAG-RNNs では，CNN と 2 次元拡張した Reccurent Neural Network(RNN) を組み合わせることで，画像全体の系列情報を捉え，局所領域と大局領域の 2 つの情報を考慮した高精度なセマンティックセグメンテーションが可能である．2D-QRNN を導入した手法は，RNN を高速化する手法である Quasi-Recurrent Neural Networks(QRNN) を用いることで，識別時間を低減することが可能となる．本研究では 2D-QRNN により高速化したセマンティックセグメンテーションを用いる．

2D-QRNN を用いたセマンティックセグメンテーションでは，QRNN の 1×2 サイズの畳み込みを 2×2 サイズに置き換えることで，2 次元に拡張する．2D-QRNN におけるメモリセルの更新は，$(+x,+y)$，$(-x,+y)$，$(+x,-y)$，$(-x,-y)$ の 4 通りの経路で行う．2D-QRNN の 4 方向の出力は，最終的に 1×1 サイズの畳み込みを行い統合する．

2D-QRNN における忘却ゲート $\mathbf{f}_t$，入力ゲート $\mathbf{i}_t$，出力ゲート $\mathbf{o}_t$ とし，算出方法を式 (1)，式 (2)，式 (3) に示す．忘却ゲートについては，軸ごとに算出する．

$$\mathbf{f}_t^d = sigmoid(\mathbf{W}_f \cdot \mathbf{a}_t + \mathbf{W}_f^d \cdot \mathbf{a}_{t-1}) \quad (1)$$

$$\mathbf{i}_t = sigmoid(\mathbf{W}_i \cdot \mathbf{a}_t + \sum_d (\mathbf{W}_i^d \cdot \mathbf{a}_{t-1})) \quad (2)$$

$$\mathbf{o}_t = sigmoid(\mathbf{W}_o \cdot \mathbf{a}_t + \sum_d (\mathbf{W}_o^d \cdot \mathbf{a}_{t-1})) \quad (3)$$

ここで，$\{d|x$ 方向, y 方向, xy の斜め方向 $\}$，$\mathbf{a}$ は前層からの入力，t は座標を示す．$\cdot$ は内積計算，$\odot$ と $+$ はユニット同士の乗算と加算である．

メモリセルへの入力候補値 $\mathbf{z}_t$ の算出方法を式 (4) に示す．

$$\mathbf{z}_t = tanh(\mathbf{W}_z \cdot \mathbf{a}_t + \sum_d (\mathbf{W}_z^d \cdot \mathbf{a}_{t-1})) \quad (4)$$

メモリセル $\mathbf{c}_t$，2D-QRNN の出力 $\mathbf{h}_t$ の算出方法を式

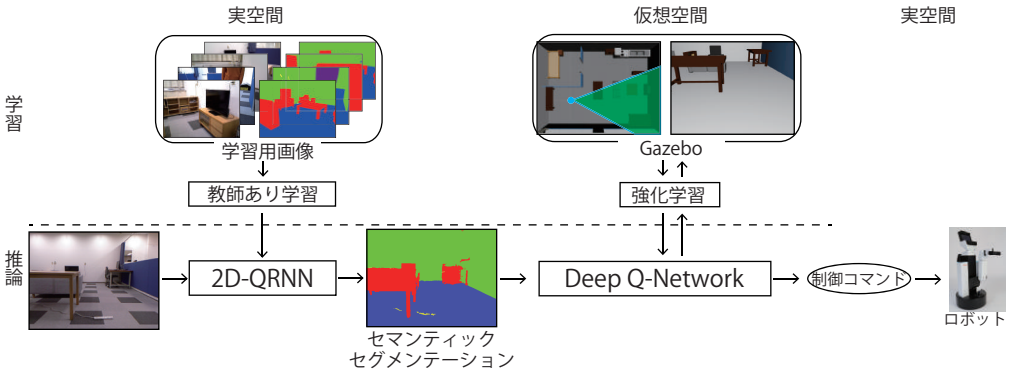


図 2: 提案するアプローチによる強化学習の流れ

(5)，式 (6) に示す．

$$\mathbf{c}_t = \sum_d (\mathbf{f}_t^d \odot \mathbf{c}_{t-1}^d) + \mathbf{i}_t \odot \mathbf{z}_t \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \mathbf{c}_t \quad (6)$$

DCNN は局所的な領域からクラスを推論するのに対して，2D-QRNN は大局的な領域からクラス推論することが可能である．従って，DCNN と 2D-QRNN を組み合わせることで，局所領域と大局領域の 2 つの情報を考慮したセマンティックセグメンテーションを行うことが可能となる．

3. 提案手法

本研究で提案するセマンティックセグメンテーションを用いた強化学習の流れを図 2 に示す．図 2 に示す学習の流れによってロボットの自律移動を学習し，実機で評価を行う．3.1 節で，2D-QRNN を用いたセマンティックセグメンテーションの学習について，3.2 節でシミュレータ上での深層強化学習での自律移動の自動獲得について，3.3 節で 2D-QRNN を用いたセマンティックセグメンテーションの結果と強化学習で獲得した行動を用いて実空間で自律移動を行う方法について述べる．

3.1 2D-QRNN によるセマンティックセグメンテーション

本研究で用いる 2D-QRNN によるセマンティックセグメンテーションのネットワーク構造を図 3 に示す．全結合層を除いた VGG-16[8] を Encoder と Decoder の構成としたネットワークをベースとする．Encoder と Decoder の間の中間層として，2D-QRNN を導入する．本研究では，室内の移動を対象とし，セマンティックセグメンテーションのクラスは，床，壁，家具，人，コード類の 5 クラスとする．リビング，ベッドルーム，ダイニングの 3 部屋で撮影した実画像をアノテーションしたものを 652 枚学習に用いた．ネットワークへの入出力画像のサイズは，横 576 画素，縦 416 画素の RGB 画像とした．図 4 に，2D-QRNN によるセマンティックセグメンテーション例を示す．

3.2 シミュレータによる Deep Q-Network による強化学習

本研究では，シミュレータを用いてロボットの自律移動のための動作を強化学習により獲得する．ある状

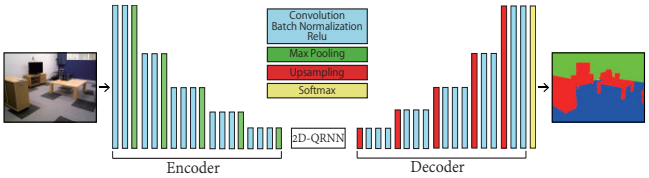


図 3: 2D-QRNN によるセマンティックセグメンテーションのネットワーク構造

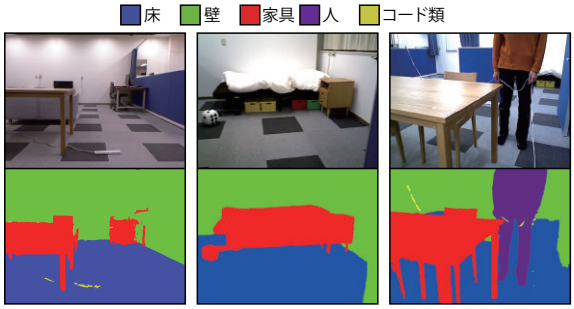


図 4: 2D-QRNN によるセマンティックセグメンテーション例

態 s の時に，エージェントが取ったある行動 a がどれだけ最適かを表す Q 値を用いた，Q 学習と呼ばれる古典的な強化学習の手法がある．Deep Q-Network（DQN）[9] は，ニューラネットワークによってその Q 値を近似する手法であり，画像のような高次元の状態空間から，行動を獲得することができる．ある状態の時に，ある行動をとることを 1 ステップとし，条件によって終了するまでのステップをまとめてエピソードと呼ぶ．割引率を γ= 0.99 とし，エピソード終了時にエージェントに対する報酬 r を用いて学習を行う．状態と行動と報酬の組み合わせを 10^5 ステップ分記憶しておき，学習時に用いる．ϵ-greedy 法を用い，確率 ϵ は 1 から 0.2 まで，1 万ステップかけて線形に減少させた．累積報酬の期待値を $\mathbb{E}$ とする．本研究で用いる，Q 値をニューラルネットワークによって近似するための損失関数 L は以下となる．

$$L(\theta) = \mathbb{E}[(r + \gamma \max_{a'} Q^*(s', a'; \theta) - Q(s, a; \theta))^2] \quad (7)$$

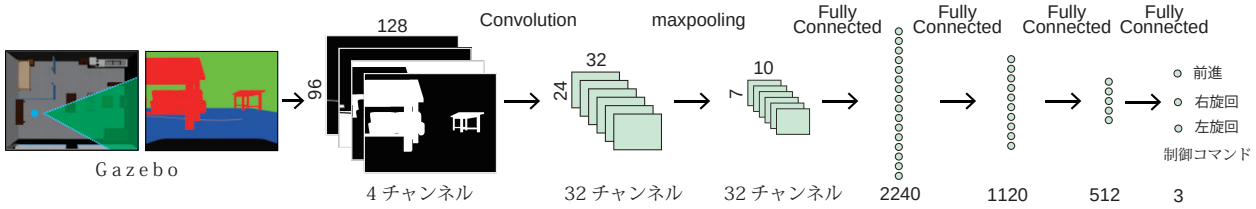


図 5: DQN ネットワーク構造

DQN に用いたネットワークを図 5 に示す．本研究では，室内の自律移動の自動獲得を対象とし，家具等の障害物がある部屋の端にゴールを設定する．シミュレータは Gazebo を用いて実際に存在する部屋（図 6(a)）のシーンからセマンティックセグメンテーションの結果を生成し，DQN への入力とする（図 6(b)）．

衝突した場合に −1 点，進入禁止エリアに進入した場合に −10 点を 1 ステップごとに与え，報酬の合計が −30 を下回るか，進入禁止エリアに進入した場合エピソードを終了する．また，ゴールに到達した場合，15 点から衝突した回数 × − 1 点を引いたものを報酬として与え，エピソードを終了する．報酬の計算式を式 (8) に示す．

$$\text{報酬} = \begin{cases} +15 - \text{衝突回数} & (\text{ゴールした}) \\ -10 - \text{衝突回数} & (\text{進入禁止に入った}) \\ -1 \times \text{衝突回数} & (\text{上記以外}) \end{cases} \quad (8)$$

エピソードが終了すると，ロボットは初期位置へと戻る．ロボットが取りうる行動は，前進，左旋回，右旋回の三つのアクションとした．

3.3 シミュレータで獲得した行動を用いた実空間での自律移動

実空間において，シミュレータで学習した結果を用いてロボットを自律移動を行う．DQN の入力は，HSR を用いて実環境で高さ 98cm の位置のカメラより取得した RGB 画像から，3.1 節で述べた 2D-QRNN によるセマンティックセグメンテーション結果を求める．次に，セマンティックセグメンテーション結果を DQN に入力し，制御コマンドである (0,1,2) を求める．制御コマンド 0 の場合は，20Hz の周期で 4 回前進の命令をロボットへ配信し，約 25cm 前進する．制御コマンド 1,2 の場合は，20Hz の周期で 3 回，1 ならば右旋回，2 ならば左旋回の命令をロボットへ配信し，約 30 度旋回する．カメラが安定しセマンティックセグメンテーションの結果が得られるまで約 2.7 秒間静止し，得られたセマンティックセグメンテーションの結果から DQN を用いて制御コマンドを受け取り，ゴールにたどり着くまで繰り返す．

4. 評価実験

4.1 実験概要

提案手法では，シミュレータ環境で学習した結果を用いて，現実空間でロボットを制御する．提案手法の評価のため，距離画像，RGB 画像でも同様に実験を行い，提案手法と比較する．まず，シミュレータ環境での学習結果の比較を行い，その後実環境でのゴール到達回数により評価を行う．

(a) 実空間　　(b) シミュレータ

図 6: 対象とする空間

4.2 シミュレータでの学習結果の比較

シミュレータを用いて DQN を 1000 エピソード学習させた際の報酬と行動回数のグラフを図 7 に示す．セマンティックセグメンテーションを入力とした場合の学習結果を図 7(a)，距離画像を図 7(b)，RGB 画像を図 7(c) に示す．どの入力を用いた場合でも，報酬の値は上昇し，行動回数である Step 数は減少していることがわかる．これは，各入力画像に対して，より少ない行動回数でゴールへと辿り着くように学習できたことを示している．

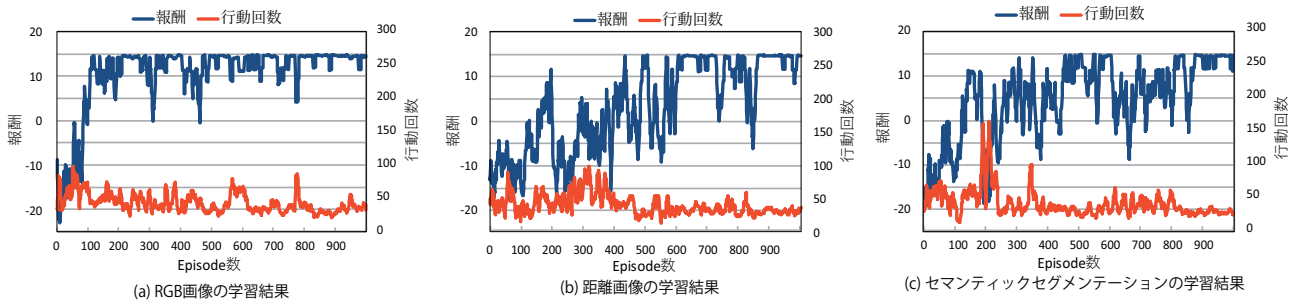
4.3 実機を用いた評価

4.2 節でシミュレータ上で学習したモデルを用いて，実空間においてロボットを 10 回動かした際の結果を表 1 に示す．ゴール地点から 1 メートルの範囲をゴールとし，障害物に衝突するか，移動せず 30 回以上行動した場合を終了条件とした．セマンティックセグメンテーションは実環境でも 10 回中 5 回のゴールに成功した．RGB 画像は 1 回も成功しなかった．距離画像では 2 回のゴールとなった．各入力におけるロボットの軌跡を図 8 に示す．ゴールまでの平均距離を比較すると，セマンティックセグメンテーションを用いた結果が良いことがわかる．これは，図 9 に示すように，RGB 画像や距離画像に比べ，セマンティックセグメンテーション画像は強化学習時の入力であるシミュレータ (図 9(d)) と実空間 (図 9(a)) で差が少ないからである．

以上より，RGB 画像や距離画像に比べ，セマンティックセグメンテーションはシミュレータでの学習結果を用いて，実世界のロボットを制御する際に有効であるといえる．

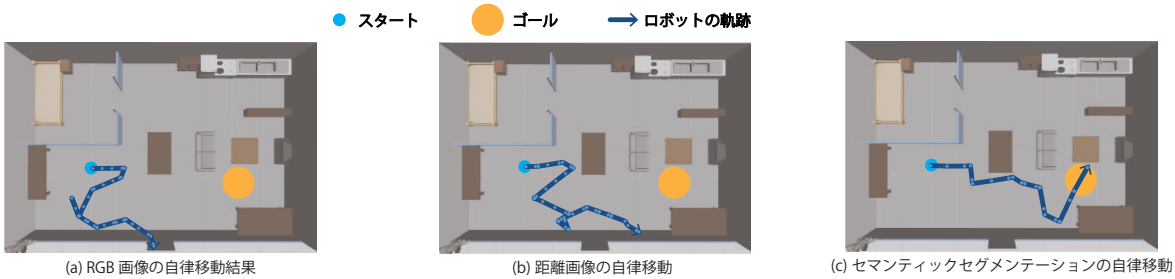
5. おわりに

本研究では，DQN での学習にセマンティックセグメンテーション結果を用いることで，実空間とシミュレー



(a) RGB画像の学習結果　(b) 距離画像の学習結果　(c) セマンティックセグメンテーションの学習結果

図 7: シミュレータでの学習結果



(a) RGB 画像の自律移動結果　(b) 距離画像の自律移動　(c) セマンティックセグメンテーションの自律移動

図 8: 実機による自律移動の例

表 1: 実験結果

入力	RGB 画像	距離 画像	セマンティック セグメンテーション
ゴール回数 [回]	0	2	5
平均距離値 [m]	4.7	1.84	1.32

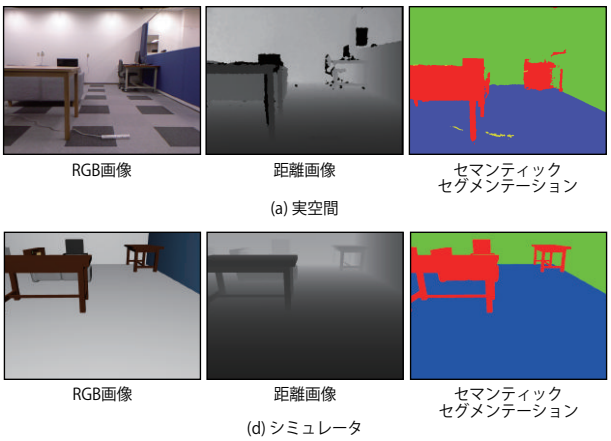


図 9: 各画像の比較

タ環境の間の差異を吸収し，DQN によりシミュレータで獲得した行動によって実空間においても行動できることを示した．今後は，より複雑な問題設定におけるセマンティックセグメンテーションを用いた強化学習に取り組む予定である．

参　考　文　献

[1] Bojarski. M, Del Testa. D, Dworakowski. D, Firner. B, Flepp. B, Goyal. P, Jackel. L, Monfort. M, Muller. U, Zhang. J, Zhang. X, Zhao. J, Zieba. L, “End to End Learning for Self-Driving Cars”, arXiv preprint:1604.07316 [cs.CV], 2016.

[2] Levine.S, Pastor.P, Krizhevsky.A, Quillen.D, “Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection”, International Symposium on Experimental Robotics, pp.173-184, 2016.

[3] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W. “OpenAI gym”, arXiv preprint arXiv:1606.01540 [cs.LG]., 2016

[4] J.Long, E.Shelhamer, and Trevor Darrell, “Fully Convolutional Networks for Semantic Segmentation”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,, pp.3431-3440, 2015.

[5] V.Badrinarayanan, A.Kendall, R.Cipolla, and S.Member, “SegNet: A Deep Convolutional EncoderDecoder Architecture for Image Segmentation”, inIEEE, arXiv preprint arXiv:1511.00561v2, 2015.

[6] B.Shuai, Z,.Zuo, G.Wang, B.Wang, “DAG-Recurrent Neural Networks For Scene Labeling”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.3620-3629, 2015.

[7] 古川 弘憲, 山下 隆義, 山内 悠嗣, 藤吉 弘亘, 石井 育規, 羽川 令子, “2D-QRNN を導入した DCNN によるセマンティックセグメンテーションの高精度化と高速化”, 画像の認識・理解シンポジウム, 2017.

[8] Simonyan.K, Zisserman.A, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, arXiv preprint arXiv:1409.1556 [cs.CV], 2014.

[9] Mnih.V, Kavukcuoglu.K, Silver.D, Graves.A, Antonoglou.I, Wierstra.D, Riedmiller.M, “Playing Atari with Deep Reinforcement Learning”, NIPS Deep Learning Workshop, 2013.