セマンティックセグメンテーションを用いた Deep Q-Network による自律移動動作の獲得

ER14071 丸山祐矢　　指導教授：藤吉弘亘

1. はじめに

ロボットの動作獲得に，深層強化学習が利用されている．深層強化学習は教師信号を必要としないが，行動を獲得するために，ロボットを実空間で何度も動作させる必要があり，学習に多大な時間を要する．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．しかし，シミュレータ環境と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学習した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．この問題を解決するために，本研究では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手法を提案する．

2. 提案手法

図 1 に提案手法の流れを示す．セマンティックセグメンテーションを中間表現として用いることで，実空間とシミュレータとの隔たりを無くすことが期待できる．これにより，シミュレータを用いて学習したロボットの自律移動動作を，ロボット実機にそのまま利用するアプローチを提案する．

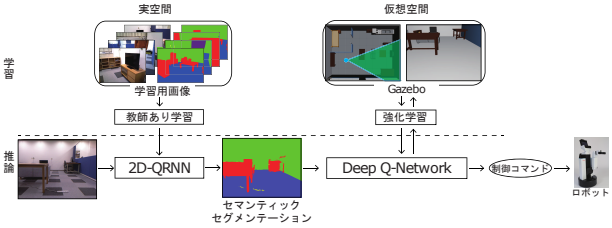

図 1：提案手法の流れ

2.1. 2D-QRNN を用いたセマンティックセグメンテーション

本研究では，2D-QRNN[1] を用いてセマンティックセグメンテーションを行う．2D-QRNN は，RNN を高速化する手法である Quasi-Recurrent Neural Networks(QRNN) を 2 次元に拡張した手法である．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．入力は，576×416 画素の RGB である．図 2 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．

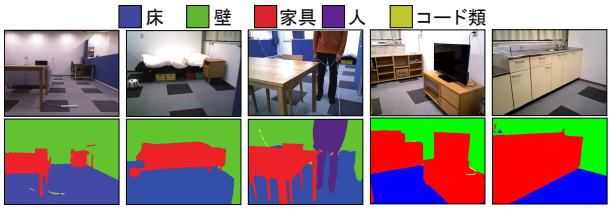

図 2：セマンティックセグメンテーション例

2.2. シミュレータによる DQN を用いた強化学習

Deep Q-Network（DQN）[2] は，ニューラルネットワークによって最適な行動を表す 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．DQN のネットワーク構成を図 3 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にランダムなスタートと固定のゴールを設定する．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する．報酬の計算を式 (1) とし，報酬の累積が −30 を下回るか，ゴールした際にエピソードを終了する．エピソードが終了すると，ロボットは初期位置へと戻る．学習はシミュレータ上で行う．

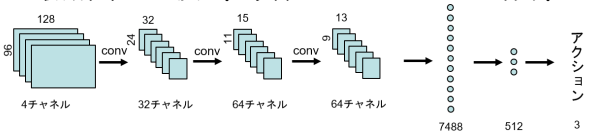

図 3：DQN のネットワーク

$$報酬 = \begin{cases} +15 - 衝突回数 & (ゴールした) \\ -1 & (衝突) \\ 0 & (上記以外) \end{cases} \quad (1)$$

2.3. 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させる．セマンティックセグメンテーション結果を DQN に入力し，制御コマンドを選択する．制御コマンドは 3 つあり，0 の場合は約 25cm 前進し，1,2 の場合は右または左に約 30 度旋回する．

3. 評価実験

提案手法での有効性を評価する．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．距離画像，RGB 画像を入力とした際の実験を行い，提案手法と比較する．

3.1. シミュレータでの学習結果の比較

セマンティックセグメンテーション，距離画像，RGB 画像を入力とし，シミュレータを用いて DQN を 2000 エピソード学習させた．報酬と行動回数のグラフを図 4 に示す．どの入力を用いた場合でも報酬が上がり，行動回数が下がり，学習ができたことを示している．

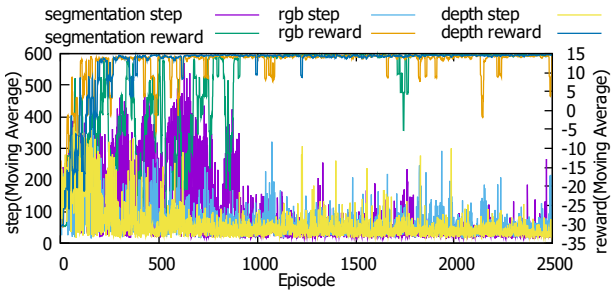

図 4：シミュレータでの学習結果

3.2. 実機を用いた評価

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．提案手法では RGB 画像，距離画像に比べ，実環境でもより多くのゴールに成功した．また，失敗時のゴールまでの平均距離を比較しても，提案手法を用いた場合の方がゴールに近くなっていることがわかる．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．

表 1：実験結果

入力	RGB 画像	距離画像	提案手法
ゴール回数 [回]	0	2	**5**
平均距離値 [m]	4.7	1.84	**1.32**

4. 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取り組む予定である．

参考文献

[1] 古川弘憲 等,“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.

[2] Mnih.V, *et al.*,“Playing Atari with Deep Reinforcement Learning”, NIPS Deep Learning Workshop, 2013.