

1. はじめに

深層強化学習は、画像のような状態数が膨大な入力に対して、畳み込みニューラルネットワークなどでの深層ネットワークを活用して、エージェントの行動を獲得することができる手法である。深層強化学習により、人が教示することなくロボットの自律移動を獲得できる [1]。しかし、その自律行動の判断根拠が明確ではないという問題がある。深層学習の判断根拠を視覚的に説明する手法が提案されており、Attention Branch Network(ABN)[2]は、推論時に着目する領域を可視化した Attention map を獲得できる。そこで、本研究では深層強化学習に Attention branch を導入し、Attention map によるエージェントの判断根拠の可視化方法を提案する。Attention map の判断根拠としての正当性の確認と、ロボットが自律移動を行う際に与える目的地の違いによる Attention map の変化を確認する。

2. Attention 機構での深層強化学習の視覚的解析

視覚的説明を目的とした手法として、ABN[2]が提案されている。ABNでは、Attention branch によって出力した Attention map を Attention 機構へ与え、注目領域を重視して識別する。しかし、ABNは教師あり学習に用いられており、強化学習では明確な教師ラベルが存在しないため、ABNを直接応用することができない。

3. 提案手法

深層強化学習では一般的に、学習コストを抑えるためにシミュレータを用いて学習することが多い。しかし、RGB画像をそのまま入力すると、ドメインギャップにより、シミュレータで学習したモデルを実空間でそのまま利用できないという問題がある。そのため、本手法ではモデルへの入力として、セマンティックセグメンテーション(Sem. Segm.)を用いる。これにより、シミュレータと実空間でのドメイン差が生じるテキストチャ情報を、ピクセル単位のラベル情報へと置き換えることで、ドメインギャップを無くすことができる。

3.1 DQN による自律移動の獲得

図1に、本研究で用いる Sem. Segm. を中間表現として用いた深層強化学習の流れを示す。学習は、シミュレータ上のテキストチャを単色化し、疑似的に Sem. Segm. 画像化する。これにより、計算コストを抑え学習を行うことができる。

深層強化学習のベース手法には、Deep Q-Network (DQN) [4]を用いる。入力としてロボット視点の画像を用い、ロボットから目的地までの距離と角度の値を、ネットワーク中間層に直接与える。距離と角度を与える方法として、図2の緑線で示すように出力層前の全結合層の出力に目的地の情報を結合する方法と、図2の赤線で示すように畳み込み層へ目的地の情報を結合する方法について比較を行う。畳み込み層へ結合する際には、角度と距離の情報をカーネルサイズ1で畳み込み処理した後、シグモイド関数の出力を得て、結合する畳み込み層のサイズに拡大する。

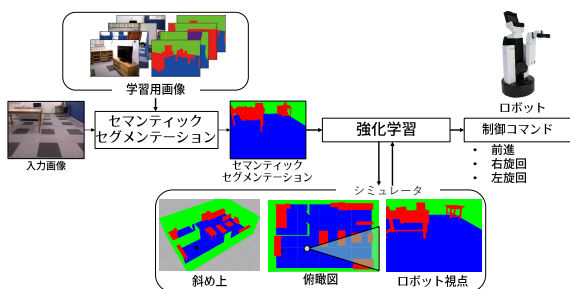


図1：深層強化学習の流れ

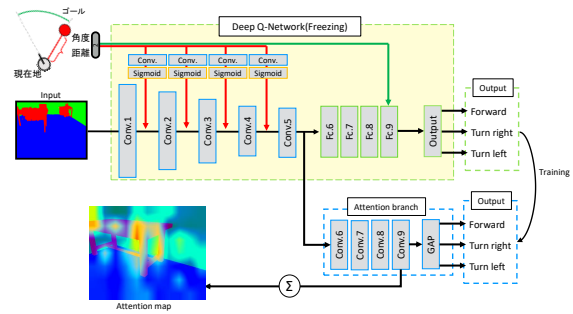


図2：本研究で用いる DQN をベースとしたネットワーク

強化学習の報酬 R は各ステップ毎に算出した r の累計とし、次式 (1) のように定義する。

$$r = \begin{cases} 30 & (\text{Reach the goal}) \\ D_c & (\text{moving forward}) \\ -5 & (\text{crash}) \end{cases} \quad (1)$$

ここで、 D_c は各ステップでのゴールまでの相対距離の変化量である。ロボットの動作は、前進、左旋回、右旋回の三種類とする。エピソードリセットの条件はゴールに到達、累計報酬 R が 30 点以下、70 ステップ以上行動の三つとし、それまでのステップをまとめて 1 エピソードとして学習、ランダムなスタートとゴールの組み合わせでリスタートを行った。

3.2 Attention branchによる視覚的判断根拠の獲得

3.1 節での深層強化学習における判断根拠を Attention map として可視化する。DQN のような Q 学習ベースの手法では、教師ラベルが存在しないため ABN を直接応用することができない。そのため、DQN で学習した学習済みモデルの重みを固定し、図2の水色枠のように Attention branch を追加する。Attention branch は畳み込み層4層と GAP からなる。固定した DQN からの出力を Attention 機構の出力に対する教師ラベルとする。教師ラベルと Attention branch の出力から Softmax Cross Entropy 誤差を求め、最小になるよう学習を行うことで Attention map を獲得する。

4. 評価実験

本実験では、ロボットの自律移動において、目的地の相対座標を全結合層へと与える方法と、畳み込み層へと与える方法を比較する。まず、deletion および insertion を用いて、Area Under Curve(AUC)を求め、獲得した Attention map が判断根拠として正しい場所を示しているかどうかを確認する。次に、Attention map を可視化し、ロボットがどのような判断根拠で行動を選択したかを視覚的に理解する。

4.1 自律移動における評価

本研究で行う深層強化学習では、ロボットが実空間の生活空間と、それをシミュレータ上で再現した空間で、自律移動するための学習と評価を行う。

実験概要

3.1 節の方法によって、約 40000 エピソード学習を行った。角度と距離の情報を、図2の緑線のとおり全結合層に相対座標を加える方法を DQN(全結合層)とし、赤線のとおり畳み込み層に相対座標を加えた方法を DQN(畳み込み層)とする。

結果

学習した結果を用いて、シミュレータで 100 回評価テストを行った結果を表1に示す。評価指標にはゴール率と、

失敗した際のゴールまでの距離 [m], および衝突回数を用いる。表 1 より, どちらのネットワークにおいても, 9 割以上のゴール率であり, 正しく学習が行われていることがわかる。

表 1: 評価結果

環境	Goal[回数]	Distance	Collisions
DQN(全結合層)	91/100	1.12	4
DQN(畳み込み層)	93/100	1.31	3

4.2 視覚的説明における評価

Attention branch は強化学習の学習と同時に学習しているわけではないため, 判断根拠として相応しい場所に Attention map が出ているとは限らない。そのため, insertion と deletion を用いて Attention map の評価を行う。

実験概要

重みを固定した学習済みの DQN に, Attention branch を追加し, 両方の出力の値が同じになるように学習を行うことで Attention map を得る。獲得した Attention map が, 判断根拠となる場所に出ているか評価する。テスト用のデータとして 1000 個の入力と, 出力のペアを用意した。入力する画像を変化させ, 正しい出力が得られるか正確さを調べる。

本研究では, Attention map に対して p タイル法を用いて二値化を行い, 入力マスクとする。また比較として, ランダムなノイズをマスクとして用いる。マスクは画像に対して 0 から 100 まで, 10%刻みで変化させる。評価は精度と, deletion および insertion を用いた Area Under Curve(AUC) によって行う。入力画像の例を図 3 に示す。

元となる画像が図 3(a) の時, 得られる Attention map は図 3(b) である。図 3(c)(d) はランダムに画像領域を deletion した画像であり, 図 3(e)(f) は Attention map に対して p タイル法を用いて 30%, 70% の二値化を行いマスクとし, deletion した画像である。図 3(g)(h) は同じく, Attention map に対して p タイル法を用いて 30%, 70% の二値化を行いマスクとし, insertion した画像である。

これらの画像を入力として用いて, 画像全体に対する deletion および insertion の割合と, テスト用データに対する Accuracy から AUC を求め評価を行う。ランダムなノイズをマスクとして用いた場合, deletion と insertion は割合を逆にしただけなので, 全体で見た AUC は変わらない。そのため, Attention map に対し deletion を行った場合とランダムに deletion を行った場合, および Attention map に対する insertion を行った場合とランダムに insertion を行った場合の 2 条件で AUC を比較する。

結果

評価結果を表 2 に示す。表 2 より, deletion において Attention の方がランダムより 0.055 低く, insertion において Attention の方がランダムよりも 0.077 高い数値であった。2.2 節の通り, deletion は低い程良く, insertion

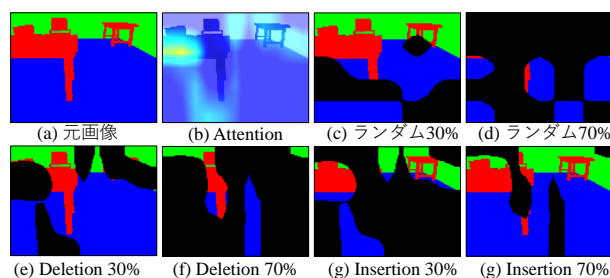


図 3: 比較画像例

は高い程良い。よって, より重要な領域に対して Attention map が得られている事が確認できた。

表 2: deletion と insertion の評価

	Attention	Random
deletion	0.596	0.651
insertion	0.728	0.651

4.3 目的地の変化による Attention map の変化

目的地の情報の与え方を変化させた時の Attention map の結果の変化を比較する。Attention map の結果例を図 4 に示す。

図 4(a)(b)(c) は, 全結合層に相対座標を加えた場合であり, 左前方から正面前方, 右前方とゴールの位置を変えた際の Attention map を示している。Attention branch に入力する段階では, ゴールの方向が与えられていないため, 全ての場合で同じ Attention map が出力されている。

それに対し, 図 4(d)(e)(f) は, Attention branch に分岐される畳み込み処理の段階で, ゴールの方向情報が与えられている。そのため, 与えた方向ごとに, その判断根拠となっている場所へ強く反応している Attention map であることがわかる。具体的には目的地を左方向とした場合, 図 4(d) のように左のソファに強く反応し, 右方向とした場合は図 4(f) のように右奥の机に強く反応している。

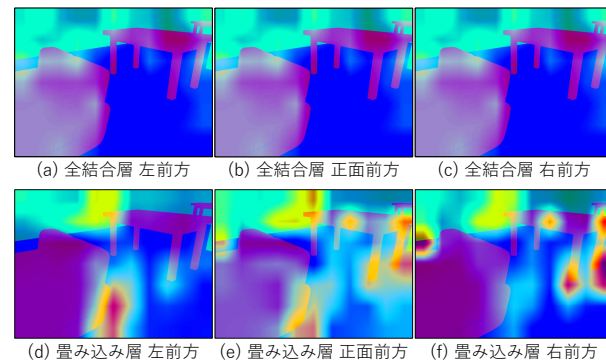


図 4: DQN(全結合) と DQN(畳み込み層) の可視化例

5. おわりに

本研究では Sem. Segm. を用いた深層強化学習による自律移動の獲得を行い, Attention branch を用いて, 判断根拠の可視化を行った。また, 獲得した Attention map が判断根拠として適しているかどうかの検証を行い, 有効であることを確認した。

参考文献

- [1] Y. F. Chen, *et al.*, “Socially aware motion planning with deep reinforcement learning”, IROS, 2017.
- [2] F. Hiroshi, *et al.*, “Attention Branch Network: Learning of Attention Mechanism for Visual Explanation”, CVPR, 2019.
- [3] V. Petsiuk, *et al.*, “RISE: Randomized Input Sampling for Explanation of Black-box Models”, BMVC, 2018.
- [4] V. Mnih, *et al.*, “Playing Atari with Deep Reinforcement Learning”, NIPS, 2013.

研究業績

- [1] 丸山祐矢 等, “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”, 日本ロボット学会学術講演会, 2017.