アピアランスベースの深層強化学習を用いたロボットマニピュレーションにおける視覚的説明

ER18063 本多航也　　指導教授：藤吉弘亘

1. はじめに

物体の把持などのロボットマニピュレーションでは，ロボットの動作を操縦者が教示するティーチングプレイバックを用いるのが一般的である．ロボットの動作環境が異なると，再度操縦者によるティーチングが必要となるため，その負担は大きい．また，ティーチング動作はロボットにとって最適であるとは限らない．そこで，深層強化学習により最適なロボット動作を獲得する手法が提案されている．しかしながら，従来の深層強化学習を用いたロボット動作は制御値を入力とするため，出力に対する判断根拠の可視化が不可能である．本研究では，画像を入力とする深層強化学習モデルに Mask-attention 機構を導入し，視覚的説明を実現する．ロボットの判断根拠を可視化することは信頼性の向上に繋がり，Human-Robot Interaction の観点からも重要である．

2. Mask-attention 機構

深層強化学習における判断根拠を可視化するために，Mask-attention A3C (Mask A3C) [1] では Mask-attention 機構を導入している．本研究では，特徴マップを注視領域でマスクすることで重要な領域を強調する処理を行う．Actor-Critic 法を用いて，Actor で action，Critic で value を出力する．Feature extractor から出力される特徴マップを入力として，畳み込み層と Sigmoid 関数を用いてアテンションマップを算出する．算出したアテンションマップを，畳み込み層と ReLU 関数で算出した action または value に対応する特徴マップにかけ合わせ，Actor または Critic の各中間層でマスク処理をする．これにより，推論時に Actor と Critic の注視領域を獲得し，最適な action と value の出力に寄与する領域を強調した学習が可能となる．

3. 提案手法

本研究はアピアランスベースの深層強化学習を用いたロボットマニピュレーションにおける視覚的説明を提案する．

3.1. ネットワーク構造

図 1 に提案手法のネットワーク構造を示す． Actor と Critic の各ブランチに Mask-attention 機構を導入し，アテンションマップを獲得する．また，現在のロボットの姿勢を考慮するために，Mask-attention 機構と全結合 (FC) 層の間にロボットの姿勢情報を入力する．

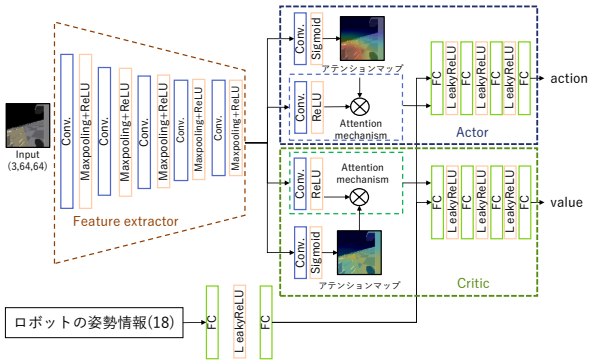

図 1：提案手法のネットワーク構造

3.2. 実験環境

本研究では，NVIDIA が提供している GPU で深層強化学習が可能なシミュレータである Isaac-gym を用い，FrankaCabinet タスクを対象とする．本タスクは，Franka Emika 社の Panda というピッキングロボットでキャビネットの引き出しを開けるタスクである．Panda とキャビネットを図 2 に示す．赤枠が Panda，黄枠がキャビネットの最上段の引き出しであり，キャビネットを開けるとタスクを達成する．姿勢情報である関節角度と関節角速度の 18 種類を入力として使用する．報酬設計を表 1 に示す．

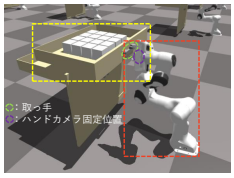

図 2：実行環境

表 1：報酬設計

報酬の条件	倍率
ハンドと引き出しの距離	×1.5
引き出しと指までの距離	×12
ハンドが引き出しに向いている	×0.5
引き出しの開き具合	×20
ハンドの右指が下，左指が上	×1.0
action 対するペナルティ	×0.01

4. 評価実験

3.2 節で説明した FrankaCabinet を対象とし，Mask-attention 機構を用いた画像入力での深層強化学習の有効性を調査する．

4.1. 実験概要

提案手法の有効性を調査するためにアテンションマップを可視化し，注視領域を確認するとともに，Mask-attention 機構の有無によるタスク達成率の変化を確認する．強化学習手法には Proximal Policy Optimization (PPO) を用いて学習する．入力画像はハンドカメラを用いて取得した，64 × 64 ピクセルの RGB 画像を用いる．学習は Isaac-gym 上で 200 台のロボットを用いて並列に 5000 エピソード行う．評価には，最も収益が高いモデルを使用する．

4.2. 実験結果

Mask-attention 機構の有無を 10 回試行した際のタスク達成率の平均を表 2 に示す．タスクの達成条件は，引き出しを全開させた場合とする．Mask-Attention 機構有りと無しでは，有りのタスク達成率が 32.75pt が向上した．このことから，Mask-attention 機構により精度が向上し，ロボットマニピュレーションに有効であると考えられる．

表 2：タスク達成率の平均の比較

Mask-attention 機構	達成率
あり	46.15%
なし	13.40%

獲得したアテンションマップを図 3 に示す．図 3 より，対象となるキャビネットに強くアテンションが発生していることから，目標物と判断したことが分かる．

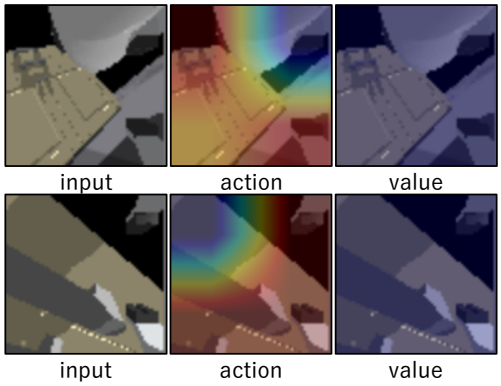

図 3：Attention map の可視化結果

5. おわりに

本研究では，画像入力の強化学習に Mask-attention 機構を導入して学習し，アテンションマップを獲得した．Mask-attention 機構を導入すると精度が向上することを確認した．今後は，画像情報を使用した強化学習の精度を向上させ，多品種のピッキングタスクを対象とする．

参考文献

[1] H. Itaya, “Visual Explanation using Attention Mechanism in Actor-Critic-based Deep Reinforcement Learning”, IJCNN, 2021.

[2] J. Schulman, *et al.*, “Proximal policy optimization algorithms”, arXiv preprint , arXiv:1707.06347, 2017.