

# 時系列深層模倣学習によるロボットアームの円軌道追従

丹羽 靖治<sup>†</sup> 相澤 宏旭<sup>‡</sup> 加藤 邦人<sup>†</sup>

<sup>†</sup>岐阜大学 工学部 〒501-1193 岐阜大学岐阜市柳戸 1-1

<sup>‡</sup>広島大学先進理工系科学研究科 〒739-8511 広島県東広島市鏡山 1-3-2

E-mail: niwa@cv.info.gifu-u.ac.jp

## <あらまし>

ロボットによる自動化を行う際、タスクの種類によっては、環境や対象に無数のパターンが考えられることがある。この場合、問題解決の筋道を確立することは難しく、ロボット設計の前段階で開発が滞ってしまう。そこで本研究では、時系列深層学習モデルを使用した模倣学習を用いて、Expert の操作からロボットに行動方策を獲得させる実験を行った。ロボットの模倣能力について定量的な評価を行うために、「ランダムな回転姿勢、中心位置、半径を持つ円軌道の追従」を簡単なタスクとして選択した。

キーワード： 模倣学習 (Imitation Learning), LSTM (Long Short-Term Memory), DART, 三角波ノイズ

## 1. はじめに

ロボット開発において、自動化するタスクの対象、環境が明確に設定できる場合、問題解決のストーリーを確立し、それに沿ってロボットを開発する。この方法はタスクの想定範囲を広げるほど、ロボットの設計や、プログラミングが複雑になる欠点を持つ。また、人間の設定した範囲からタスクのパターンが外れると、ロボットが正しく動作する保障はない。人間が日常で容易に処理するタスクの中には、環境や対象に無数のパターンが存在するタスクもある。これらに対し、具体的な問題解決の筋道を確立することは困難である。よって全く別のアプローチで自動化を図る必要がある。

人間の遠隔操作情報からロボットに方策を獲得させる手法に、模倣学習<sup>[1]</sup> (Imitation Learning)がある。Expert の操作履歴から Agent の方策を学習する本手法は、現実の様々なタスクの自動化と、ロボットの設計の単純化につながる。本研究では、時系列深層学習モデルを使用して人間の行動データからロボットに方策を獲得させる基盤作りを行う。実験では、ランダムな回転姿勢、中心位置、半径を持つ円に沿って、アームの先端を追従させるための方策をロボットに獲得させる。

## 2. シミュレータ

実験で使用するシミュレータについて説明する。シミュレータの全体像を図 1 に示す。図 2 のようにロボットアームの先端には半径 5.0[mm] の赤い制御球が取り付けられている。ロボットは制御球の中心座標が目標点の座標と重なるように逆運動学で動く。

ロボットの前方には円が描かれた板を設置する。板はランダムな回転姿勢を持つ。板に描かれる円は、中心位置と半径がランダムで決定される。制御球は円周上の角度 0 の位置からスタートする。板とロボットは物理的に干渉せず、円は固定される。円軌道のランダムな範囲を表 1 に示す。

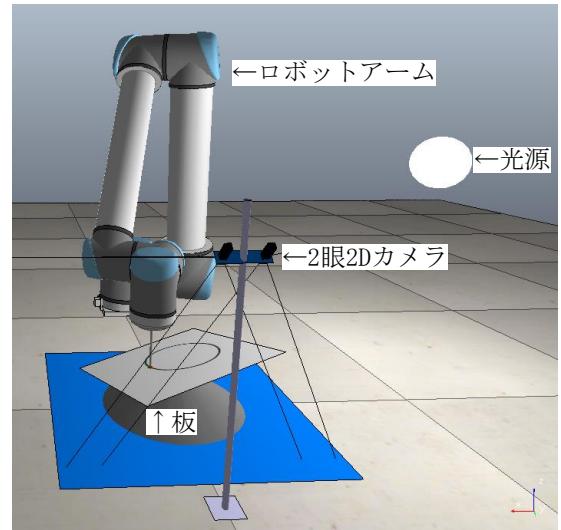


図 1 シミュレータの全体像

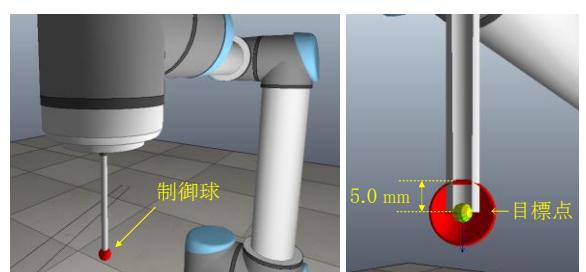


図 2 ロボットの制御球と目標点

次に図 3 の 2 眼 2D カメラについて説明する。図 1 のようにテクスチャボードを見下ろす位置に 2 眼カメラが取り付けられ、三人称視点でロボットと板を観測する。カメラから得られる左右画像を操縦者の左右の目に VR ゴーグルを介して提示することで、シミュレータ空間を立体的に認識させる。カメラから得られる左右画像の例を図 4 に示す。

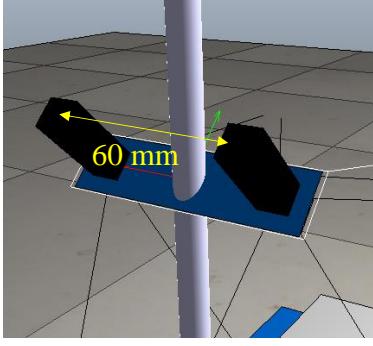


図 3 2 眼 2D カメラ

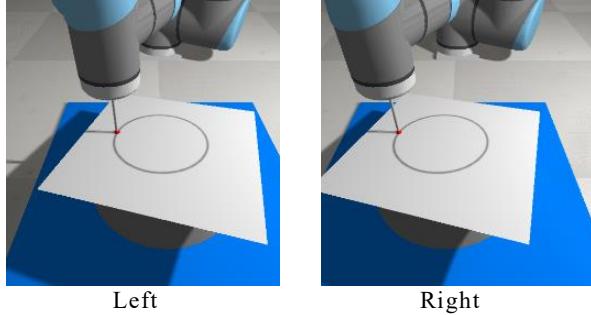


図 4 2 眼 2D カメラから得られる左右画像の例

表 1 円のランダムパラメータ

板の回転角 $(\alpha, \beta, \gamma)$ [deg]	$\alpha$ ( $x$ 軸回転) [−15, 15]
	$\beta$ ( $y$ 軸回転) [−15, 15]
	$\gamma$ ( $z$ 軸回転) [0, 360]
半径 [px]	[70, 100]
縦横方向シフト [px]	[−30, 30]
シフト前中心座標 [px]	(152, 152)

$x, y, z$  はシミュレータの世界座標系を表す。円の中心座標の縦横シフトは、円のテクスチャ座標系を用いる。

### 3. 目的

実験の目的について説明する。定められた範囲でランダムな円に対し、制御球中心と円の線が重なるように反時計回りで正確に追従する方策をロボットに獲得させることを本実験の目的とする。

### 4. 手法

本研究では模倣学習を使用し、ロボットに方策を獲得させる。模倣学習は Expert の状態と行動の組から、Agent の方策を学習する手法である<sup>[1]</sup>。本実験において、Expert は人間の遠隔操縦者、Agent はシミュレータ上のロボットになる。模倣学習では、状態を方策に入力して出力される予測行動と、Expert の行動との誤差が小さくなるようにパラメータを最適化する。実験では Expert の操作系に三角波ノイズ<sup>[2]</sup>を 3 次元的に付与することで理想的な軌道から外れたときの復帰行動を学習させる。図 5 に操作ベクトルのローカル座標系の  $x_{local}$  軸の回転方向に対し、実際に付与された三角波ノイズ  $\alpha_{noise}$  [deg] の一部を示す。図 5 中の  $t$  は時間ステップを表し、1 ステップは 0.25 秒の時間幅を持つ。

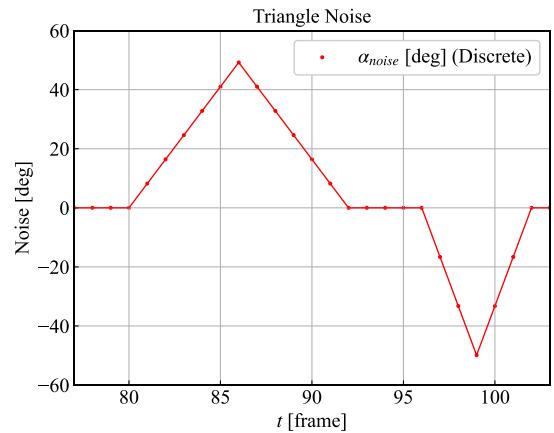


図 5 三角波ノイズ  $\alpha_{noise}$

### 5. データセット

200 パターンの 3 次元円に対し、Expert は目的にしたがって、4fps 間隔で 200 回制御球を移動させる。この過程で左右画像と操作量を連番で保存する。約 2.7 時間分のデータを収集し実験を行う。

### 6. ネットワーク

実験では 2 つのネットワークを使用し比較する。どちらのネットワークも 2 眼 2D カメラの左右画像から、相対的な移動を行うための操作量を予測する。

1 つ目の V3BPM (Vanilla 3D Behavior Prediction Model、図 6) は、LSTM<sup>[3]</sup>を使用した時系列モデルの前段に、画像 2 枚から特徴抽出を行うための 3 つのエンコーダを接続したネットワークである。左右画像を個別に処理する単眼エンコーダ (Monocular

Encoder)と左右のパスから得られた特徴マップをチャネル方向に連結し、畳み込む両眼エンコーダ(Binocular Encoder)で3次元空間認識を行う。

2つ目のネットワーク C3BPM (CNN<sup>2</sup> 3D Behavior Prediction Model, 図7) は単眼エンコーダを CNN<sup>2</sup> [4]

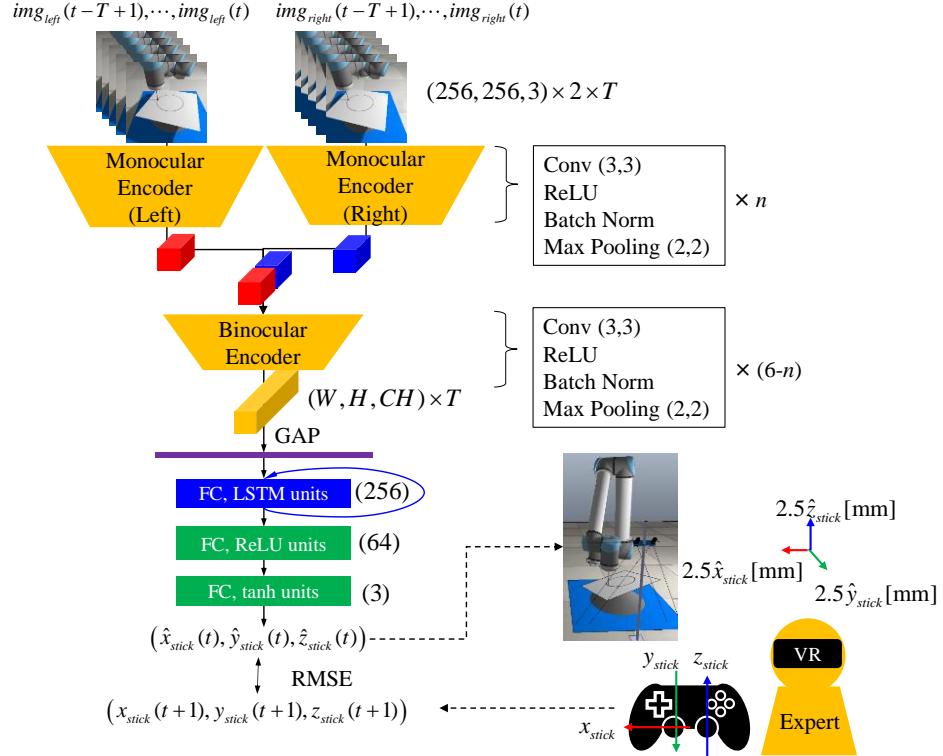


図6 Vanilla 3D Behavior Prediction Model

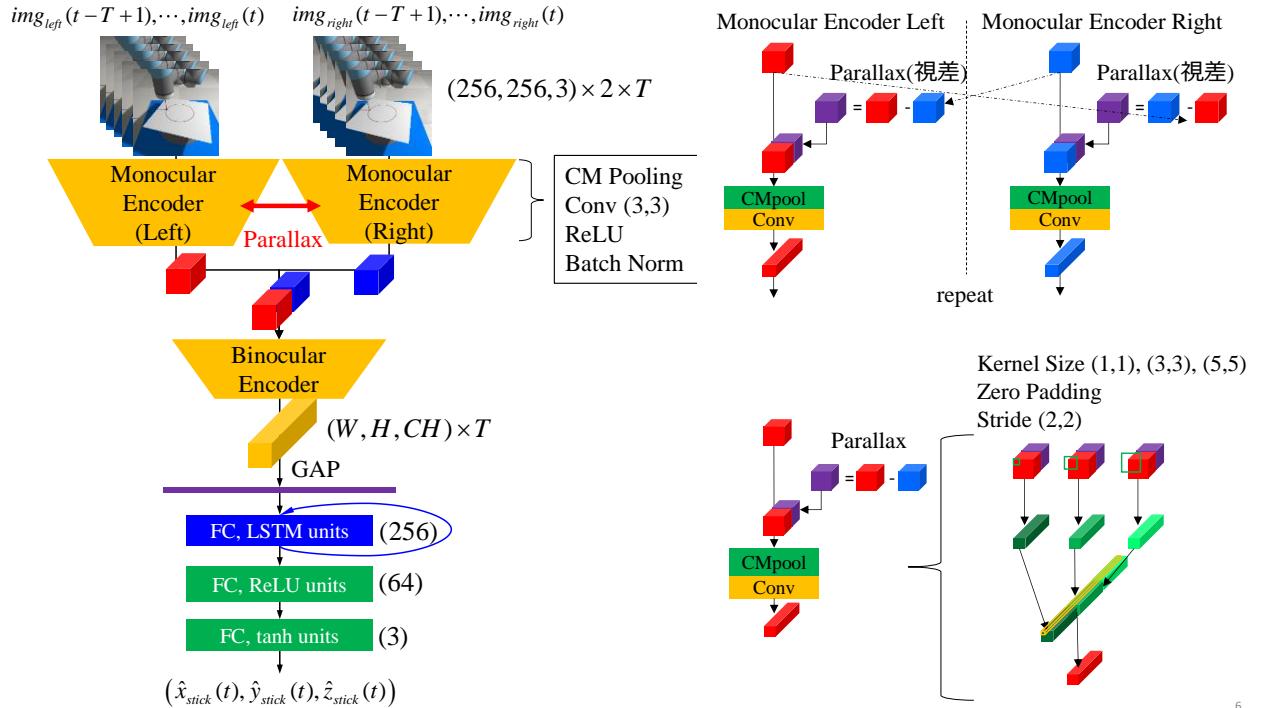


図7 CNN<sup>2</sup> 3D Behavior Prediction Model

に置き換えたモデルである。CNN<sup>2</sup>部で左右の単眼エンコーダのパスに視差情報、焦点ずれを明示的に組み込むことで、視差による両眼立体視と、焦点のずれによる単眼立体視を行うモデルとなっている。

## 7. 実行性能の評価方法

ノイズなし環境下における、Expert と Agent の操作精度を比較する。図 8 のように、制御球中心と円の最短点との距離  $d$  と、最短点が円周上を進む速度  $v$  の 2 点で追従精度を定量的に評価する。 $v$  は反時計回りを正とする。Expert および各 Agent に、学習に使用していない 15 個の円を与え、各円で 200 回移動を行ない、合計 3000 個の  $d$  と  $v$  を得る。この分布でモデルの追従精度を評価する。

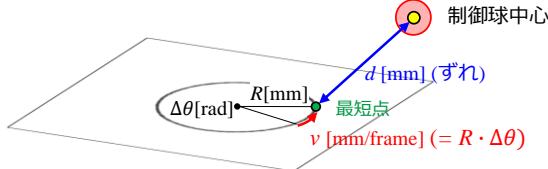


図 8 評価値  $d$  と  $v$

## 8. V3MPM と C3BPM の比較

V3MPM と C3BPM の実行性能を比較する。系列長  $T=5$  のデータセットを使用して訓練した各モデルの結果を表 2 に示す。表 2 から、 $v$  に関しては大きな違いは見られなかった一方で  $d$  の最大値、すなわち最も円から遠ざかったときの  $d$  を比較すると、C3BPM の方が V3BPM よりも小さな値をとる傾向があり、C3BPM の追従精度が高いことが分かる。

## 9. 時系列、非時系列、Expert の比較

時系列モデル、非時系列モデルのベストチューニングモデルと Expert の操作精度を比較する。図 9 にそれぞれの  $d$  の分布を示す。また、図 10 にテストコース 163 番の人間と時系列モデルの描いた軌跡を示す。

図 9 の  $d$  の平均値は、Expert が最も小さな値となり、平均的な追従精度が最も高いことが分かる。次点は時系列モデルで、わずかに非時系列モデルより良い結果を示した。一方  $d$  の最大値に関しては、時系列モデルが最も小さな値をとり、Expert のようなうつかりミスが起こりにくいといえる。図 10 からも Expert の操作より時系列モデルの方がより滑らかな

軌道を描いたことが分かる。これらから Expert の平均的な操作をモデルが学習したことが分かる。

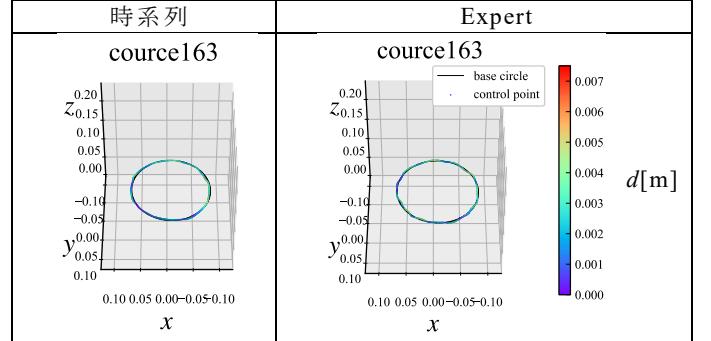


図 10 人間とロボットの軌跡比較

## 10. まとめ

人間の遠隔操作履歴からシミュレータ上のロボットにランダムな円を追従する方策を獲得させた。本研究で築いた基盤を活用し、今後は Peg-in-Hole など高度で実用的なタスクの自動化を図る。また、より直観的な遠隔操作を行うために、Oculus Quest 2 などの市販の VR 機器を遠隔操作デバイスとすることも今後の課題とする。

## 文 献

- [1] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne: “Imitation Learning: A Survey of Learning Methods”, ACM Computing Surveys, Vol. 5, No. 21, (2017)
- [2] M. Laskey, J. Lee, R. Fox, A. Dragan, and K. Goldberg: “DART: Noise Injection for Robust Imitation Learning”, arViv:1703.09327v2 [cs.LG], (2017)
- [3] S. Hochreiter, and J. Schmidhuber: “LONG SHORT-TERM MEMORY”, Neural Computation, Vol. 9, No. 8, 1735-1780, (1997)
- [4] W. Chen, and S. Wu: “CNN<sup>2</sup>: Viewpoint Generalization via a Binocular Vision”, NeurIPS 2019, (2019), <http://www.cs.nthu.edu.tw/~shwu/pubs/shwu-neurips-19.pdf>

表 2 V3BPM と C3BPM の畠み込み層数の変更

モデル	比	$d$			$v$		
		mean	std	max	mean	std	min
V3BPM	1:5	5.50	3.90	21.68	2.65	0.505	-0.036
	2:4	4.77	4.01	27.88	2.60	0.467	-0.031
	3:3	3.97	2.72	19.87	2.61	0.450	-0.036
	4:2	5.21	4.16	29.81	2.60	0.474	-0.030
	5:1	4.55	3.13	16.59	2.64	0.392	-0.040
C3BPM	1:5	5.94	4.69	28.41	2.56	0.504	-0.031
	2:4	4.35	2.95	16.98	2.64	0.463	-0.037
	3:3	3.18	1.53	9.16	2.59	0.437	-0.036
	4:2	<b>2.96</b>	<b>1.36</b>	<b>8.26</b>	2.58	0.417	-0.036
	5:1	3.12	1.55	9.57	2.63	0.422	-0.032

比は単眼エンコーダと両眼エンコーダの畠み込み層数の比を表す。  
mean は平均値、std は標準偏差、max は最大値、min は最小値を表す。

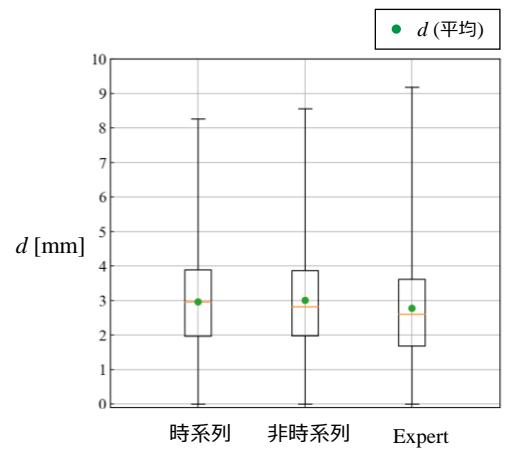


図 9  $d$  の箱ひげ図