

# 強化学習を用いたドローン点検経路の自動生成

## Automatic Generation of Drone Inspection Routes by Reinforcement Learning

○柴田拓海<sup>\*1</sup>, 山田悟史<sup>\*2</sup>

Takumi SHIBATA<sup>\*1</sup>, Satoshi YAMADA<sup>\*2</sup>

\*1 立命館大学 大学院 理工学研究科 博士課程前期課程

Graduate Student, Graduate School of Science and Engineering, Ritsumeikan Univ.

\*2 立命館大学 理工学部 建築都市デザイン学科 准教授 (工学)

Associate Professor, Dept. of Architecture and Urban Design, Ritsumeikan Univ., Dr.Eng.

キーワード: ドローン ; 維持管理 ; 点検 ; 強化学習 ; 自動生成

Keywords: Drone ; Maintenance ; Inspection ; Reinforcement learning ; Automatic generation.

### 1. はじめに

#### 1.1 社会課題・背景

今日、建設業界において、人手不足やインフラの維持管理の問題を抱えている。日本の社会インフラである橋梁やトンネルの多くは高度経済成長期に集中的に建設された。その結果、多くの構造物は老朽化の目安といわれる50年を超え、今後急速的に老朽化が進む懸念がある。国土交通省<sup>1)</sup>によると建設後50年を超える構造物は、2023年3月に道路橋で39%、トンネルで27%になると予測されている。(表1)。それに対して定期的な点検による維持管理が必要不可欠といえる。また、国土交通省は5年に1度の点検を義務付けた。そのような点検には専門技術者が不足しており、維持管理における大きな課題となっている。特に自治体においては、技術系職員の不足やインフラ老朽化に対する財源不足を抱えている。

これに対し、近年発達している情報技術の活用が求められている。例えば、IoT化したインフラ状況の常時点検やAIによる画像診断、ドローンを用いた点検の省力化などがあげられる。中でも特に人工知能、AI、の発展が著しく、インフラの維持管理での活用を期待されている。カメラによる画像から機械学習を用いて、コンクリートの点検をし、ひび割れなどの異常を発見するという研究が多くなされている<sup>2)3)</sup>。また、建設業界でのドローンの活用も広まっている。ドローンでの撮影により、3次元モデルを構築し、現場での安全管理に用いることや障害物を避けながらの被害状況の撮影、さらには資材の運搬といった試みがなされている<sup>4)</sup>。ドローンの点検は従来に比べ費用を抑えることができるが、操縦技術者の不足を背景に普及が進んでいない。

#### 1.2 研究目的

そこで本研究では、ドローンでの構造物点検の経路生成に関する研究を行った。建設業界における人手不足と

表1(建設後50年を超える構造物の割合)

	2018.3	2023.3	2033.3
道路橋(約37万橋 長2m以上の橋)	約25%	約39%	約63%
トンネル(約1万1千本)	約20%	約27%	約42%
河川管理施設(約1万施設)	約32%	約42%	約62%
下水道管(総距離:約4万km)	約4%	約8%	約21%
港湾岸壁(約5千施設,水深4.5以深)	約17%	約32%	約58%

インフラの維持管理の必要性といった課題をドローン操作を自動化することによって改善できると考えた。様々な建物の形状や敷地の条件に応じて、ドローンを操縦するのは高い技術力が必要であり、技術者が現場に行くには時間的・日数的に限りがある。そこで、構造物の壁面のドローン点検検査を想定して、飛行経路を強化学習により生成する手法について検討した。ここでは、複数エージェントでの最適化を目標とし、基礎段階として単独エージェントでの強化学習による最適化を図った。単独ドローンを強化学習により制御し、自動で壁面検査を行うまでの流れを示すことを目標とした。強化学習を用いて、自動化した例として、ドローンの配送経路の最適化やクレーン揚重、車の駐車自動化といったものがある<sup>5)6)</sup>。しかし、ドローンによる壁面検査の飛行経路の強化学習は、筆者が知る限り見当たらず、この部分が新規性と言える。そこで、本研究ではゲームエンジンによる強化学習を用いて、簡略化したドローンの壁面点検経路の自動生成を行った。壁面は2種類用意した。1つ目は1面のみの壁面、2つ目は「く」の字型の2面の壁面である。

### 2. 研究概要

#### 2.1 学習環境

強化学習<sup>7)</sup>は機械学習の1つで、環境の状態と報酬に基づいて、最良の行動が取る方策を学習する手法である。図2に示すように強化学習は次の①~③の手順を繰り返

すことで学習する。

- ① エージェントは時間ステップ  $t$  ごとに状態を取得
- ② 取得結果に基づき、行動を選択
- ③ 選択した行動により、報酬が与えられる

以上の、手順を繰り返しながら、長期的な収益値を最大化する行動を見つけていく。

本研究では、状態・行動・報酬の設計、学習・学習結果の確認を Unity を用いて行った。強化学習は Unity 向けの機械学習ツールキットである Unity Machine Learning Agent (ML-Agents v2.0.1)<sup>8)</sup> を用いた。強化学習のアルゴリズムには学習アルゴリズムの 1 つであるオフポリシーの Proximal Policy Optimization (PPO) 法<sup>9)</sup> を使用した。PPO はサンプル効率が低く、学習に必要なステップ数は多いが、学習が安定しているという特徴がある。また、学習に必要なハイパーパラメータの調整は個別に行った。

## 2.2 環境設計

ドローンでの壁面検査をゲームエンジンの Unity によって再現した (図 3)。ドローンと壁面は学習を効率化するために簡略化した。ドローンではモデルと動き方を簡略化した。ドローンのモデルは立方体で表現し、動き方は、上下、左右、前後、左右への回旋に制限した。また、壁面は 1 面の壁を複数の小さな正方形 (マス) の組み合わせで構成した。

3次元空間にこの簡略化した壁面を設置し、ドローンが壁面検査をするよう設計した。壁面検査は次のような状況になることで検査をしたとみなすこととした。「ドローンのカメラを壁面に対して正面を向ける」かつ「ドローンと壁面との距離が一定以内になった時」の状況である。これは、高性能カメラの発達により、カメラが壁面をとらえることができたなら、検査は十分に行えるということに基づいて設定した。シュミレーション環境ではドローンから一定距離にレイを放った。レイが壁面に接した際に、当たった壁面の一部の正方形の色をグレーから赤に変化させた (図 3)。赤になった部分は検査したとみなし、疑似的にカメラによる壁面検査を再現した。

## 3. 一面壁面における経路生成

1つの壁面のみを作成し、ドローン検査の経路を強化学習させた。壁面のマスを検査した時と全ての壁面の検査をし終えた時に報酬を与えるように設計した。ドローンが範囲外に出てしまった時と 1面全ての検知が完了した時にそれぞれ罰則と報酬を与え、エピソードを終了させ、初期位置から再スタートさせた。壁の位置とドローンの行動できる範囲を図 4 と表 2 に示した。ドローンの行動できる範囲が狭いと罰則を嫌って中央に留まる傾向があったため表のような行動範囲に設定した。さらに、ドローンが停滞することなく、効率よく点検を完了させるように時間経過ごとの罰則も与えた。また、その他の強

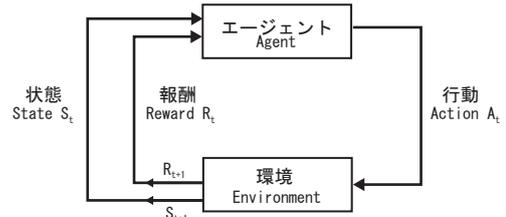


図 2 (強化学習の概略)

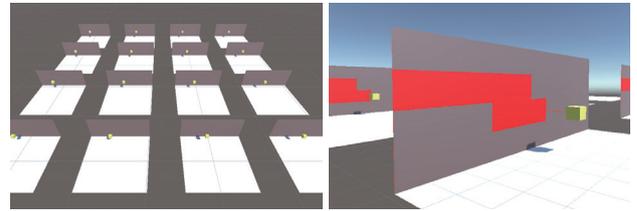


図 3 (壁面とドローンの学習環境・検知の様子)

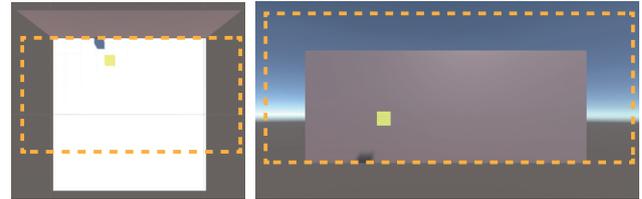


図 4 (ドローンの行動可能範囲)

表 2 (壁の位置とドローンの行動可能範囲)

	壁の位置座標	ドローンの行動可能範囲
x 範囲	$-5 \leq x \leq 5$	$-8 \leq x_d \leq 8$
y 範囲	$0 \leq y \leq 4$	$0 \leq y_d \leq 8$
z 範囲	$z = 5$	$-4 \leq z_d \leq 5$

表 3 (設定 1 の環境設計)

行動 Action	ドローン上下左右前後移動 (x, y, z 座標移動)	
状態観察 Observation	ドローン x, y, z 座標	3 observations
報酬 Reward	1 マスを検知	+1/1000
	全ての壁面を検知し終わり	+1.0
	範囲外	-1/1000
	毎時	-1/10000

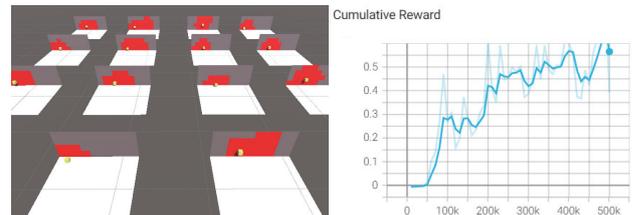


図 5 (3.1 の検知の様子・平均累積報酬)

化学習する際の状態の観察方法・報酬の与え方・報酬量は、それぞれの学習で個別に設計した。

### 3.1 設定 1 (ドローンの停滞)

#### 3.1.1 環境設計の概要

設定 1 ではドローンの動きを上下・左右・前後の  $x, y, z$  座標移動の 3 つのみとした。状態の観察にはドローンの  $x, y, z$  座標を取得した。報酬では 1 マスの検知・すべてを検知完了時にプラスの報酬、範囲外に出た時・毎時にマイナスの報酬を与えた。詳細の数値に関しては表 3 に示した。エピソード開始時にはドローンが壁の前のランダ

ムな位置に配置されるようにした。また、ハイパーパラメータの1つである beta を 5/10000 とした。

### 3.1.2 学習結果

ドローンが図5の状態でも中央に留まり続け、壁面の検知が行われないという結果となった。この時の学習による500Kエピソードまでの学習の平均累積報酬のグラフを図5に示した。毎時の罰則を大きくした際には探索することなくすぐに範囲外に出てしまっていた。

## 3.2 設定2(未検査のマスが残る)

### 3.2.1 概要

設定2では、未検査壁面の検知を ml-agent のコンポーネントである Raycast を用いて行った。これは物体から ray を放ち特定の物体に当たった際にその物体の位置を特定できるものである。ここでは、ドローンから ray を放ち、検査未実施の壁面のみを検知するように設定した。ray の本数・角度に関しては図7に示した。また、ドローンの行動として y 軸回転を加えた。その他の環境設計について表4にまとめた。エピソード開始時のドローンの位置は3.1と同様ランダムな位置とした。また、ハイパーパラメータの beta を 1/100 とし、行動決定のランダムさを増加させ探索を活発にさせた。

### 3.2.2 学習結果

円を描くように壁面を検知するようになった。しかし、1部分に検査の残しが発生し、検査済みの壁面の前を移動し続けてしまうことになった。この時のエピソード中における時系列でのドローンの検査経路の1例を図6に示した。最短経路での壁面全体の検知はできなかった。また、学習による平均累積報酬のグラフを図7に示した。

## 3.3 設定3(全ての面の検査可能)

### 3.3.1 概要

設定3では、壁面の位置を特定するために、Raycast に加えてドローンと壁面までの距離とベクトルを追加で取得するようにした。この際、エピソード中に検知した壁面の情報は観察から外し、未検知の壁面のみを取得した。また、水平にドローンを誘導しようと考え、水平方向にあるマスを検知したときは、それ以外のマスを検知した時の報酬に比べ大きくなるようにした。その時、検知した前回のマスと現在のマスの距離が近いほど報酬が大きくなるように設計した。図8に検知したマスの距離とドローンの移動方向に応じた報酬量の違いを図示した。また、エピソード開始時のドローンの位置は左上の壁面の前に固定した。その他、詳細を表5に示した。

### 3.3.2 学習結果

ドローンは車体を左右に振りながら水平の壁面を検査しつつ、上下方向に動くことで全ての壁面を検知していた。エピソード中における時系列でのドローンの検査経路の1例を図9に示した。また、学習による平均累積報酬とエピソードの長さのグラフを図10に示した。

表4(設定2の環境設計)

行動 Action	ドローン上下左右前後移動 (x, y, z 座標移動)	
	ドローン旋回 (y 軸まわりの回転)	
状態観察 Observation	ドローン x, y, z 座標	3observations
	ドローン y 軸回転角度	1observation
	RayCast Observation	5rays
報酬 Reward	1 マスを検知	+1/1000
	全ての壁面を検知し終わり	+1.0
	範囲外	-1/1000
	毎時	-1/10000

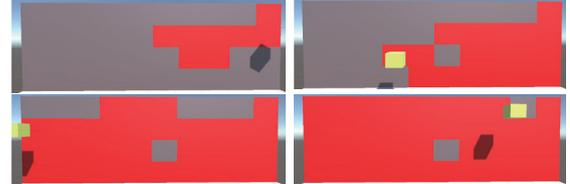


図6(時系列での検査状況)

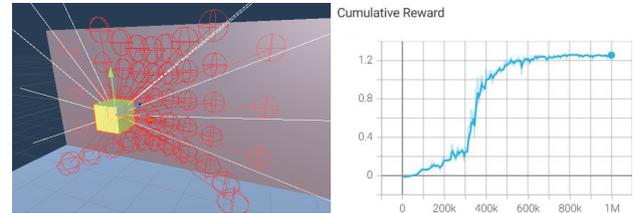


図7(3.2のRaycastによる観察・平均累積報酬)

表5(設定3の環境設計)

行動 Action	ドローン上下左右前後移動 (x, y, z 座標移動)	
	ドローン旋回 (y 軸まわりの回転)	
状態観察 Observation	ドローン x, y, z 座標	3observations
	ドローン y 軸回転角度	1 observation
	ドローンとマスまでの距離	0~40
	ドローンからマスまでのベクトル	0~120
	RayCast Observation	5rays
報酬 Reward	垂直にある壁の検知をした時	+1/100*distance*
	それ以外の壁面を検知した時	+1/1000*distance*
	全ての壁面を検知し終わった時	+1.0
	範囲外	-1/1000
	毎時	-1/1000000

\* distance は1つ前に検知したマスと現在検知したマスの距離を示す(最小値=1.0)

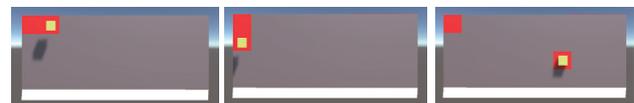


図8(報酬量, 左:大, 中央:中, 右:小)



図9(時系列での検査状況)

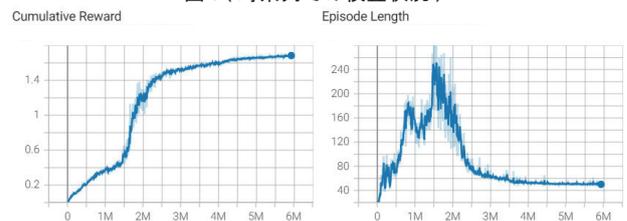


図10(3.3の平均累積報酬とエピソードの長さ)

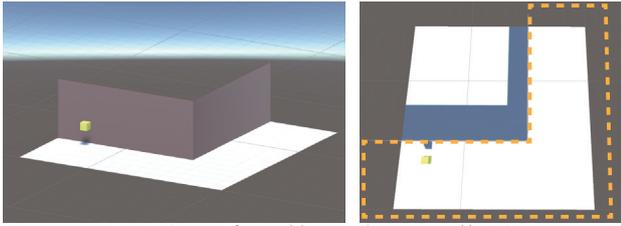


図 11(二面壁面の様子とドローンの範囲)

表 6(壁の位置とドローンの行動範囲)

壁の座標	壁の座標(側面)	行動範囲 ( $x_d < 5$ のとき)	行動範囲 ( $5 \leq x_d$ のとき)
$-5 \leq x \leq 5$	$x=5$	$-5 \leq x_d \leq 5$	$5 < x_d \leq 12$
$0 \leq y \leq 4$	$0 \leq y \leq 8$	$0 \leq y_d \leq 8$	$0 \leq y_d \leq 8$
$z=5$	$5 \leq z \leq 15$	$-2 \leq z_d \leq 5$	$-2 \leq z_d \leq 18$

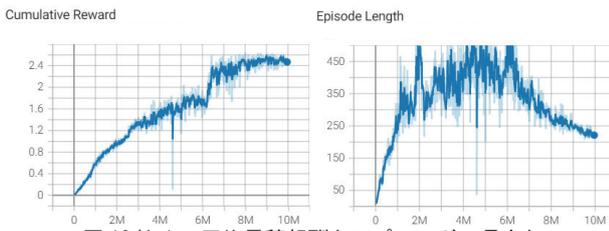


図 12(4.1 の平均累積報酬とエピソードの長さ)

#### 4. 二面壁面における経路生成

2つの壁面を図 11 のように「く」の字型に配置し、同様に学習を行った。壁面の位置とドローンの行動できる範囲を図 12 と表 6 に示した。状態・報酬・行動の設計は一面壁面の時と同様に個別に設計した。

##### 4.1 設定 3(全ての面の検査可能)

###### 4.1.1 概要

学習環境は設定 3 と同様の設計とした(表 5)。エピソード開始時のドローンの位置は左上の壁面の前に固定した。

###### 4.1.2 学習結果

この時の学習による平均累積報酬とエピソードの長さのグラフを図 12 に示した。水平に移動しながら壁面全体の検知が可能となった。しかし、水平移動時に検知の残しが発生してしまっていた。そのため側面を検知した後、再び前面に戻って残りの壁面を検知しており、最善の検知経路とは言えず課題となった。

##### 4.2 設定 4(全ての面の検査可能)

###### 4.2.1 概要

設定 4 では、ドローンを垂直に誘導するような報酬を与えた。すなわち、設定 3 では水平方向に移動し検査した場合に大きな報酬としたが、設定 4 では垂直移動時に大きな報酬を付与するようにした。それ以外の設定については設定 3(表 5) と同様にした。

###### 4.2.2 学習結果

正面の壁面から垂直に移動しながら検知しつつ横に移動していき、全ての面を検知可能となった。しかし、垂直移動の際に、一番下の面まで検知することなく横に移動してしまっていた。そのため、4.1 と同様に側面の検知後に再び前面に戻って残りの壁面を検知しており、最善

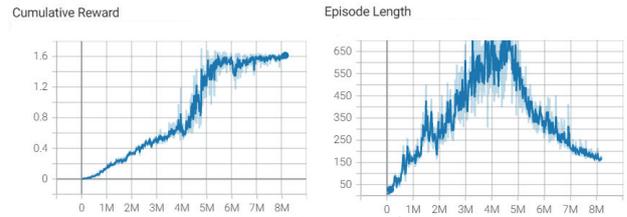


図 13(4.2 の平均累積報酬とエピソードの長さ)

の検知経路とは言えず課題となった。平均累積報酬とエピソードの長さのグラフを図 13 に示した。垂直誘導した場合、4.1 の学習方法に比べ、収束までの学習量が少なかった。さらに、全ての壁面を完了させるまでのエピソード時間も 4.1 よりも短い結果となった。

#### 5. まとめ

強化学習による壁面検査の経路自動生成における、状態観察・報酬・行動の設計の検証をした。1面壁面も2面壁も同様に、Raycast 観察と壁面の距離とベクトルによる観察をし、水平に誘導するような報酬の与え方をしたところ、全ての壁面を検知できるようになった。2面では垂直誘導時に、全ての壁面を完了させるまでの時間が短くなった。しかし、全ての壁面を検知はできてはいるが効率的なドローンの移動経路とはなっておらず、状態観察・報酬・行動やハイパーパラメータの再設計が課題となった。

#### [参考文献]

- 国土交通省, “社会資本の老朽化の現状と将来”, [https://www.mlit.go.jp/sogoseisaku/maintenance/02research/02\\_01.html](https://www.mlit.go.jp/sogoseisaku/maintenance/02research/02_01.html), (参照 2022-09-28)
- 山根 達郎, 全 邦 釘: Deep learning による Semantic Segmentation を用いたコンクリート表面ひび割れの検出, 構造工学論文集 A, 2019, 65A 巻, p. 130-138
- 中村 秀明, 山本 拓海, 青島 亘佐: コンクリート道路橋の点検省力化に向けた変状領域の自動抽出, インフラメンテナンス実践研究論文集, 2022, 1 巻, 1 号, p. 386-393
- 日経クロステック, “全自動ドローンが“現場監督”に、日々の出来高測量や安全巡視を無人化”, 2021.08.03 [https://xtech.nikkei.com/atcl/nxt/column/18/00107/00171/?i\\_cid=nbpnxt\\_reco\\_atype](https://xtech.nikkei.com/atcl/nxt/column/18/00107/00171/?i_cid=nbpnxt_reco_atype), (参照 2022-09-28)
- 樽谷 葵, 石田 航星: 深層強化学習を用いた揚重計画の自動作成手法に関する研究—2 点間の揚重経路作成における衝突回避の実装—, 日本建築学会関東支部研究報告集, 91 号, p. 417-420, 2021-03
- Guillem Muñoz, Cristina Barrado, Ender Çetin, Esther Salami: Deep Reinforcement Learning for Drone Delivery, Drones 2019, 3(3), 72; <https://doi.org/10.3390/drones3030072>
- Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore.: Reinforcement learning: A survey. Journal of artificial intelligence research, 4:237-285, 1996.
- Arthur Juliani, Vincent-Pierre Berges, Ervin Teng, Andrew Cohen, Jonathan Harper, Chris Elion, Chris Goy, Yuan Gao, Hunter Henry, Marwan Mattar, Danny Lange, : Unity: A General Platform for Intelligent Agents, arXiv:1809.02627, <https://doi.org/10.48550/arXiv.1809.02627> Focus to learn more
- John Schulman, Filip Wolski, Prafulla Dhari-wal, Alec Radford, and Oleg Klimov.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347, 2017.