Deep Q Reinforcement Learning for Autonomous Navigation of Surgical Snake Robot in Confined Spaces

S. Athiniotis, R. A. Srivatsan and H. Choset

Robotics Institute, Carnegie Mellon University,

sathinio@andrew.cmu.edu

INTRODUCTION

Airway management is fundamental for all anesthetic as well as emergency medicine procedures to maintain airway patency, prevent aspiration and permit ventilation without leakage. While endotracheal intubation and tracheostomy are regarded as the go-to procedures in such incidents, they are reportedly correlated with numerous side effects which can sometimes even be life-threatening [1]. These complications stem from the fact that essentially a human is blindly and manually maneuvering the intubation tube. In order to mitigate the ensued risks and aftereffects of currently employed methods, this work uses a surgical snake robot [2] to autonomously navigate down the airway.

The contribution of this paper is developing the navigation policy that utilizes images from a monocular camera mounted on its tip. We use Q Reinforcement Learning in Deep Convolutional Neural Networks (DCNN) [3], widely referred to as Deep Q Reinforcement Learning Neural Networks (DQNN), to produce these policies. The system can serve as an assistive device for medical personnel to perform endoscopic intubation, with minimal to no human input.

MATERIALS AND METHODS

Reinforcement learning (RL) has proven to be a successful tool for autonomous navigation and control over the past years. The objective of RL is to learn good policies for sequential decision problems by optimizing accumulated reward indicators [4]. In particular, learning to control agents from sensory data like vision, had been of great interest to the community [5, 6], including advancements introduced by Google DeepMind [3]. Most methods employ DCNN in Q learning implementations of RL algorithms. These innovative approaches empower agents to accurately predict optimum courses of thousands of sequential actions, in the presence of noisy input data in dynamic environments.

Although deep RL is currently facing an unprecedented resonance in the majority of engineering fields, it has not yet been used widely in the medical domain. To the best of our knowledge, no existing literature suggests its deployment for endoscopic intubation. This paper demonstrates that a DQNN framework can accurately enable a flexible snake robot to navigate inside a patient's airway using camera images (Fig 1 (b)).

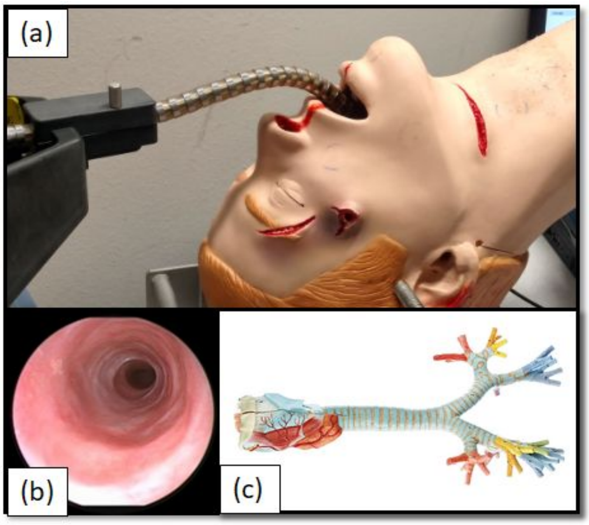

Fig. 1 (a) Surgical snake robot in assistive respiration. (b) Image of trachea as seen from the camera on the head of the robot. (c) The trachea is a non uniform curved tube of length 100-120 mm and diameter 18-20 mm.

Given the inter-patient variability as well as poor lighting and featureless environment, conventional motion planning and computer vision algorithms do not produce results with high levels of confidence. However, data driven RL algorithms rely solely on information from the unknown Markovian[1] environment. Hence, eliminating the prevalent need of most machine learning methods for large labeled datasets. Therefore we consider RL.

In this work, a surgical snake robot (see Fig. 1 (a)) plays the role of the agent[2]. The underline control algorithm interprets its state by propagating the camera's live feed through the DCNN and predicting the optimum action. In order to train the network in a realistic context, we formulated a Gazebo simulation as shown in Fig 2. Since the robot is a follow-the-leader mechanism [2], we restrict our learning to the actions of the tip of the robot alone. This helps reduce the computational complexity of the model and significantly accelerates the training time.

The sensory input supplied to the system are 84×84×3 RGB images from a camera mounted at the tip of the robot's head which carries its own light source.

[1] Markovian is used to describe a fully observable environment where the state transition probability function depends on the future state given the present state [4].

[2] In RL, agent is the entity that learns how to perform the intended task (for example robot, vehicle, etc.) [4].

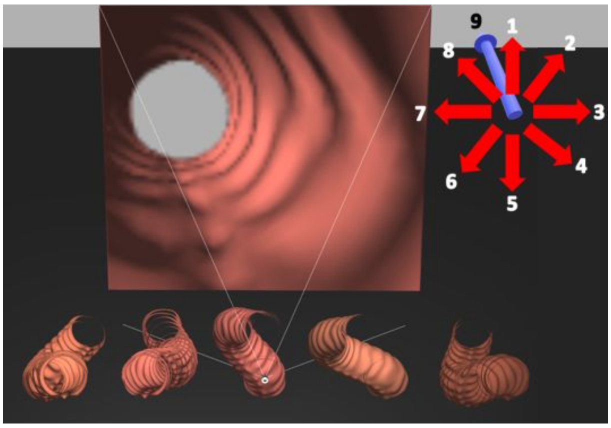

Fig. 2 Gazebo training environment including the agent (snake scope head), five different trachea models on which the DCNN was trained, the projection of the camera frame and the 9 discrete actions the agent is allowed to take.

The DCNN has 4 convolutional hidden layers followed by 2 fully connected layers which are split into two to form the proposed dueling structure [6]. The number of outputs is equal to the discrete number of possible actions that the agent is allowed to take, which in our case is 9 (See Fig. 2). In each of the defined actions, the head of the snake is oriented to the desired pitch/yaw by 5 deg and then advanced one step in the forward direction by 0.5 mm.

The network is iteratively trained for 80,000 episodes using stochastic gradient descent to update the weights over a batch of 32 experiences. The exploration rate, which reflects to the randomness of the selected actions, is linearly decreased from 100% to 1% after 70,000 episodes (see Fig. 3). The reward function is defined such that the further the distance traversed, higher the assigned reward. The termination conditions for each episode are - (1) Reaching the end of the trachea, (2) completing more steps than allowed, and (3) colliding with the walls. In order to detect collision between the solid bodies, two 2D lasers were mounted on the coronal and transverse plane of the agent. However, they will not be included in the actual experiment and they are only used for ease of collision detection in simulation.

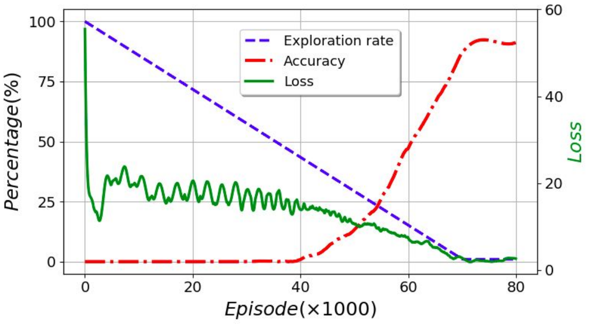

Fig. 3 Accuracy and loss versus number of training episodes.

In order to improve the robustness to different environments, and smoothly transfer to real world scenarios, domain normalization [7] is implemented by:

- Adding Gaussian noise to the camera data
- Varying the color of each model
- Randomizing agent's initial position and pose
- Randomly selecting a trachea model every 10 episodes

The following hyperparameters for the learning algorithm were hand-tuned through iterative trainings: discount rate=0.98, learning rate=0.0005, collision penalty=40, completion reward=100, buffer memory size=60,000 and batch size=32.

RESULTS

In Fig. 3 we present the resulting accuracy over training episodes, which eventually converged to 92%. In our setup, accuracy is calculated as the number of successes in 100 consecutive episodes, where success is counted only if the agent reaches the end of the trachea without meeting any of the other two terminating conditions [3]. We also present the mean loss of each episode of our training in Fig. 3, which is calculated as the average of the quadratic loss values over all steps of an episode.

DISCUSSION

We have implemented an autonomously navigating agent within a trachea using DQNN. This is challenging because the trachea is highly confined and thus two consecutively wrong actions could have inevitably lead to unwanted collision, i.e. failure. To address this challenge, we implemented a reinforcement learning approach, which produces policies that only select the appropriate action to navigate. We believe the approach in this paper can serve as an assistive device for a broad spectrum of endoscopic procedures. Of particular interest to the authors involve extending this approach to natural orifice transluminal surgery (NOTES) and endoscopic submucosal dissection (ESD).

REFERENCES

[1] J. Divatia and K. Bhowmick. Complications of endotracheal intubation and other airway management procedures. *Indian J Anaesth*, 49(4):308-318, 2005.

[2] A. Degani, H. Choset, A. Wolf, and M. A. Zenati. Highly articulated robotic probe for minimally invasive surgery. In *Proceedings ICRA,* 2006. (pp. 4167-4172).

[3] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra and M. Riedmiller. Playing Atari with Deep Reinforcement Learning. *arXiv preprint arXiv:1312.5602,* 2013.

[4] R. S. Sutton and A. G. Barto. Reinforcement learning: An introduction. *MIT press,* 2018.

[5] V. H. Hado, A. Guez, and D. Silver. Deep Reinforcement Learning with Double Q-Learning. In 30th *AAAI Conference on Artificial Intelligence*, 2016.

[6] Z. Wang, T. Schaul, M. Hessel, H. V. Hasselt, M. Lanctot, and N. De Freitas. Dueling Network Architectures for Deep Reinforcement Learning. *arXiv preprint arXiv:1511.06581,* 2016.

[7] V. H. Hado, A. Guez, M. Hessel, V. Mnih and D. Silver. Learning values across many orders of magnitude. *arXiv preprint arXiv:1602.07714,* 2016.

[3] See experimental result in this Video link