Multiple Landmark Detection Using Multi-agent Reinforcement Learning

Athanasios Vlontzos(✉), Amir Alansary, Konstantinos Kamnitsas, Daniel Rueckert, and Bernhard Kainz

BioMedIA, Computing Department, Imperial College London, London, UK
athanasios.vlontzos14@imperial.ac.uk

Abstract. The detection of anatomical landmarks is a vital step for medical image analysis and applications for diagnosis, interpretation and guidance. Manual annotation of landmarks is a tedious process that requires domain-specific expertise and introduces inter-observer variability. This paper proposes a new detection approach for multiple landmarks based on multi-agent reinforcement learning. Our hypothesis is that the position of all anatomical landmarks is interdependent and non-random within the human anatomy, thus finding one landmark can help to deduce the location of others. Using a Deep Q-Network (DQN) architecture we construct an environment and agent with implicit inter-communication such that we can accommodate K agents acting and learning simultaneously, while they attempt to detect K different landmarks. During training the agents collaborate by sharing their accumulated knowledge for a collective gain. We compare our approach with state-of-the-art architectures and achieve significantly better accuracy by reducing the detection error by 50%, while requiring fewer computational resources and time to train compared to the naïve approach of training K agents separately. Code and visualizations available: https://github.com/thanosvlo/MARL-for-Anatomical-Landmark-Detection

1 Introduction

The exact localization of anatomical landmarks in medical images is a crucial requirement for many clinical applications such as image registration and segmentation as well as computer-aided diagnosis and interventions. For example, for the planning of cardiac interventions it is necessary to identify standardized planes of the heart, *e.g.* short-axis and 2/4-chamber views [1]. It also plays a crucial role for prenatal fetal screening, where it is used to estimate biometric measurements like fetal growth rate to identify pathological development [17]. Moreover, the mid-sagittal plane, commonly used for brain image registration and assessing anomalies, is identified based on landmarks such as the Anterior

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-32251-9_29) contains supplementary material, which is available to authorized users.

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11767, pp. 262–270, 2019.
https://doi.org/10.1007/978-3-030-32251-9_29

Commissure (AC) and Posterior Commissure (PC) [2]. Manual annotation of landmarks is often a time consuming and tedious task that requires significant expertise about the anatomy and suffers from inter- and intra-observer errors. Automatic methods on the other hand can be challenging to design because of the large variability in the appearance and shape of different organs, varying image qualities and artefacts. Thus, there is a need for methods that can learn how to locate landmarks with highest accuracy and robustness; one promising approach is based on the use Reinforcement Learning (RL) algorithms [2,8].

Contributions: This work presents a novel Multi-Agent Reinforcement Learning (MARL) approach for detecting multiple landmarks efficiently and simultaneously by sharing the agents' experience. The main contributions can be summarized as: *(i)* We introduce a novel formulation for the problem of multiple landmark detection in a MARL framework; *(ii)* A novel collaborative deep Q-network (DQN) is proposed for training using implicit communication between the agents; *(iii)* Extensive evaluations on different datasets and comparisons with recently published methods are provided (decision forests, Convolutional Neural Networks (CNNs), and single-agent RL).

Related Work. In the literature, automatic landmark detection approaches have adopted machine learning algorithms to learn combined appearance and image-based models, for example using regression forests [16] and statistical shape priors [6]. Zheng et al. [19] proposed using two CNNs for landmark detection; the first network learns the search path by extracting candidate locations, and the second learns to recognize landmarks by classifying candidate image patches. Li et al. [13] presented a patch-based iterative CNN to detect individual or multiple landmarks simultaneously. Ghesu et al. [8] introduced a single deep RL agent to navigate in a 3D image towards a target landmark. The artificial agent learns to search and detect landmarks efficiently in an RL scenario. This search can be performed using fixed or multi-scale step strategies [7]. Recently, Alansary et al. [2] proposed the use of different Deep Q-Network (DQN) architectures for landmark detection with novel hierarchical action steps. The agent learns an optimal policy to navigate using sequential action steps in a 3D image (environment) from any starting point towards the target landmark. In [2] the reported experiments have shown that such an approach can achieve state-of-the-art results for the detection of multiple landmarks from different datasets and imaging modalities. However, this approach was designed to learn a single agent for each landmark separately. In [2] it has also been shown that performance of different strategies and architectures strongly depends on the anatomical location of the target landmark. Thus we hypothesize that sharing information while attempting simultaneous detection reduces the aforementioned dependency.

Background: Reinforcement Learning (RL) allows artificial agents to learn complex tasks by interacting with an environment E using a set of actions A. The agent learns to take an action a at every step (in a state s) towards the target solution guided by a reward signal r during training. The main goal is to maximize the expected rewards in order to find the optimal policy π^*. In Q-Learning, a state-action value function $Q(s, a)$ is used to approximate the value of taking an action in a given state. The Q-function is defined as the expected value of the accumulated discounted future rewards, which can be approximated iteratively as: $Q_{t+1}(s, a) = E[r + \gamma \text{max}_a(Q_t(s', a'))]$. Here $\gamma \in [0, 1]$ is a discount factor that is used to incorporate the notion of uncertainty in future events. Mnih et al. [15] proposed an approximation of the Q-function using a CNN by optimizing the network cost $L(\theta) = E\left[\left(r + \gamma \max_{a'} Q_{target}(s', a'; \theta^-) - Q_{net}(s, a; \theta)\right)^2\right]$. Q_{target} is a temporary fixed version of Q_{net}, which gets updated every N_{target} steps, used in order to avoid destabilization caused by rapid policy changes.

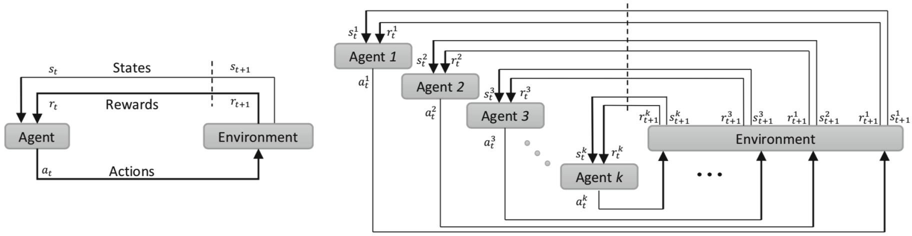
In single-agent RL scenarios, individual models learn solely from states that result from the actions of an agent. Complementary to this, MARL models learn from states that result from multiple agents dynamically interacting with their shared environment. In MARL models, there are K agents interacting with environment E. Each learns to take an action a_t^k during a state s_t^k using a reward signal r_t^k. Thus, the environment is subjected to the actions of all agents, as shown in Fig. 1. Hence, the environment becomes non-stationary as action a_i in state s_k will not always lead to the same future, since the future state is also a function of the other agents. This causes a violation of the Markov assumptions needed for the formulation of a RL scenario as a Markov Decision Process (MDP). To address this issue, [5,18] proposed to establish communication between the agents, thus taking all agents actions into account.

Any agent communication signifies the exchange of information or knowledge about the underlying Markov state of the environment. Communication between agents can be achieved explicitly via a communication protocol like in [4], where a limited bandwidth channel is learned by the agents, or implicitly by sharing knowledge in the parameter space or by combining value functions [10]. MARL scenarios can be classified as collaborative or competitive depending on the relation of the communication between agents. In this paper, we define the collaborative scenario as agents that attempt to minimize a common loss function. Competition between agents signifies a scenario in which agents try to minimize their own loss function through increasing the loss function of other agents.

2 Proposed Method

In this work, we formulate the problem of multiple anatomical landmark detection as a multi-agent reinforcement learning scenario. Building upon the work of [2,8] we extend the formulation of landmark detection as a Markov Decision Process (MDP), where artificial agents learn optimal policies towards their target landmarks, which defines a concurrent Partially Observable Markov Decision Process (co-POMDP) [9]. We consider our framework concurrent as the agents train together but each learns its own individual policy, mapping its private observations to a personal action [10]. We hypothesize that this is necessary as the localization of different landmarks requires learning partly heterogeneous policies. This would not be possible with the application of a centralized learning system.

Our RL framework is defined by the *States* of the environment, the *Actions* of the agent, their *Reward Function* and the *Terminal State*. We consider the environment to be a 3D scan of the human anatomy and define a state as a Region of Interest (ROI) centred around the location of the agent. This makes our formulation a POMDP as the agents can only see a subset of the environment [11]. We define the frame history to be comprised of four ROIs. In this setup each agent can move along the x, y, z axis creating thus a set of six actions. The agents evaluate their chosen actions based on the maximization of the rewards received from the environment. The reward function is defined as the relative improvement in Euclidean distance between their location at time t and the target landmark location. In our multi-agent framework, each agent calculates its individual reward as their policies are disjoint.



(a) Single Agent

(b) Multi-Agents

Fig. 1. (a) A single agent and (b) multi agents interact within an RL environment.

During training, we consider the search to have converged when the agent reaches a region within 1 mm of the target landmark. Episodic play is introduced in both training and testing. In training, the episode is defined as the time the agents need to find the landmarks or until they have completed a predefined maximum number of steps. In case one agent finds its landmark before all others, we freeze the training and disable network updates derived from this agent while allowing the other agents to continue exploring the environment. During testing, we terminate the episode when the agent starts to oscillate around a position or exceeds a defined maximum number of frames seen in the episode similar to [2].

Collaborative Agents. Previous approaches to the problem of landmark detection by [2,7,8] considered a single agent looking for a single landmark. This means that further landmarks needs to be trained with separate instances of the agent making a large scale application unfeasible. Our hypothesis is that the position of all anatomical landmarks is interdependent and non-random within the human anatomy, thus finding one landmark can help to deduce the location of other landmarks. This knowledge is not exploited when using isolated agents. Thus, in order to reduce the computational load in locating multiple landmarks and increase accuracy through anatomical interdependence, we propose a collaborative multi agent landmark detection framework (Collab-DQN). The following description will assume just two agents for simplicity of presentation. However, our approach scales up to K agents. For our experiments we show evaluations using two, three and five agents trained together.

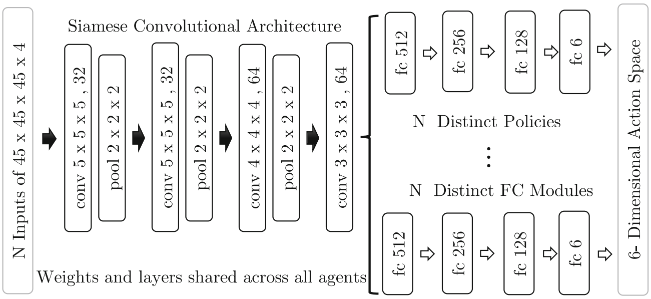


Fig. 2. Proposed Collaborative DQN for the case of two agents; The `convolutional` layers and corresponding weights are shared across all agents making them part of a Siamese architecture, while the policy making fully connected layers are separate for each agent

A DQN is composed of three `convolutional` layers interleaved with `maxpool` layers followed by three `fully connected` layers. Inspired by Siamese architectures [3], in our Collab-DQN we build K DQN networks with the difference that weights are shared across the `convolutional` layers. The `fully connected` layers remain independent since these will make the ultimate action decisions constituting the policy for each agent. In this way, the information needed to navigate through the environment are encoded into the shared layers while landmark specific information remain in the fully connected ones. In Fig. 2, we graphically represent the proposed architecture for two agents. Sharing the weights across the `convolutional` layers helps the network to learn more generalized features that can fit both inputs while adding an implicit regularization to the parameters avoiding overfitting. The shared weights enable indirect knowledge transfer in the parameter space between the agents, thus, we can consider this model as a special case of collaborative learning [10].

3 Experimentation

Dataset: We evaluate our proposed framework and model on three tasks: (i) brain MRI landmark detection with 728 training and 104 testing volumes [12]; (ii) cardiac MRI landmark detection with 364 training and 91 testing volumes [14] and (iii) landmark detection in fetal brain ultrasound with 51 training and 21 testing volumes. Each modality includes 7–14 anatomical ground truth landmark locations annotated by expert clinicians [2].

Training: During training an initial random location is chosen from the inner 80% of the volume, in order to avoid sampling outside a meaningful area. The initial ROI is $45 \times 45 \times 45$ pixels around the randomly chosen point. The agents follow an ϵ-greedy exploration strategy, where every few steps they choose a random action from a uniform distribution while during the remaining steps they act greedily. Episodic learning with the addition of freezing action updates for the agents that have reached their terminal state until the end of the episode is used, as detailed in Sect. 2.

Table 1. Results in millimeters for the various architectures on landmarks across brain MRI and fetal brain US. Our proposed Collab DQN performs better in all cases except the CSP where we match the performance of the single agent.

Method	AC	PC	RC	LC	CSP
Supervised CNN	–	–	–	–	5.47 ± 4.23
DQN	2.46 ± 1.44	2.05 ± 1.14	3.37 ± 1.54	3.25 ± 1.59	$\mathbf{3.66 \pm 2.11}$
Collab DQN	$\mathbf{0.93 \pm 0.18}$	$\mathbf{1.05 \pm 0.25}$	$\mathbf{2.52 \pm 2.25}$	$\mathbf{2.41 \pm 1.52}$	3.78 ± 5.55

Testing: For each agent, we fixed 19 different starting points in order to have a fair comparison among the different approaches. These points were used for all testing volumes for each modality at 25%, 50% and 75% of the volume's size. For each volume the Euclidean distance between the end location and the target location was averaged for each agent for each of the 19 runs. The mean distance in mm was considered to be the performance of the agent in the specific volume.

Multiple tests have been performed using our proposed architecture. Comparisons are made against the performance on multi-scale RL landmark detection [7], fully supervised deep Convolutional Neural Networks (CNN) [13] as well as a single agent DQN landmark detection algorithm [2]. In case of cardiac landmarks we compare with [16] that utilizes decision forests. Different DQN variations like the Double DQN or Duelling DQNs are not evaluated since their performance provides little to no improvement for the task of anatomical landmark detection as exhibited in [2].

Even though our method can scale up to K agents given enough computational power we limited our comparison to the Anterior Commissure (AC) and the Posterior Commissure (PC) of the brain; the Apex (AP) and Mitral Valve Centre (MV) of the heart; the Right Cerebellum (RC), Left Cerebellum (LC) and Cavum Septum Pellucidum (CSP) for the fetal brain. These are common, diagnostically valuable landmarks used in the clinical practice and by previous automatic landmark detection algorithms. For completeness and to facilitate future comparisons, we provide our performance comparison also for the training of three and five agents simultaneously. In Table 1, we show the performance of the brain MRI and fetal brain US landmarks using the different approaches. In Table 2 we exhibit the results for three and five agents trained simultaneously and the results for cardiac MRI landmarks.

Discussion: As shown in Tables 1 and 2 our proposed method significantly outperforms the current state-of-the-art in landmark detection. p-values from a paried student-t test for all experiments were in the range 0.01 to 0.0001. We perform an ablation study by training instances of a single agent with double the iterations and double the batch size. The study has been conducted on the Cardiac MRI landmarks that have exhibited the biggest localization difficulties because of larger anatomical variations across subjects than observed in brain data. Our results confirm that the agents share basic information across them, which helps all of them perform their tasks more efficiently. These results support our hypothesis that the regularization effect from the gradients collected from the increased experience and knowledge of the multi-agent system is advantageous. Furthermore, we created a single agent with doubled memory but due to the random initialization of experience memory, the agent failed to learn. In addition, as shown in Table 2(a), the inclusion of more agents leads to similar or improved results across all landmarks. It is interesting to note that even though we perform better in all landmarks, our approach can only match the performance of a single agent DQN for the CSP landmark. We theorize that this is due to the different anatomical nature of the RC, LC landmarks compared to the CSP landmark, thus the joint detection does not present an advantage. We chose to utilize the DQN in this paper rather than existing policy gradient methods like A3C as the DQN is represented by a single deep CNN that interacts with a single environment. A3C use many instances of the agent that interact asynchronously and in parallel. Multiple A3C agents with multiple incarnations of such environments are computationally expensive. In future work, we will investigate the application of other methods for multiple-landmarks detection using either collaborative or competitive agents.

Computational Performance: Training multiple agents together does not only provide benefits in performance of landmark localization, it also reduces the time and memory requirements of training. Sharing the weights between the convolutional layers helps to reduce the trainable parameters by 5% in case of

Table 2. (a) Multiple agent performance, training and testing were conducted in the Brain MRI; Landmarks 3, 4, 5 represent respectively the outer aspect, the inferior tip and the inner aspect of the splenium of corpus callosum; (b) multi-agent performance on cardiac MRI dataset;

Landmark	3 Agents	5 Agents
AC	0.94 ± 0.17	0.98 ± 0.25
PC	0.96 ± 0.20	0.90 ± 0.18
Landmark 3	1.45 ± 0.51	1.39 ± 0.45
Landmark 4	N/A	1.42 ± 0.90
Landmark 5	N/A	1.72 ± 0.61

(a)

Method	AP	MV
Inter-Obs. Error	5.79 ± 3.28	5.30 ± 2.98
Decision Forest	6.74 ± 4.12	6.32 ± 3.95
DQN	4.47 ± 2.64	5.73 ± 4.16
DQN Batch ×2	4.30 ± 12.07	5.01 ± 4.49
DQN Iterations ×2	4.78 ± 13.87	5.70 ± 18.11
Collab DQN	$\mathbf{3.96 \pm 5.07}$	$\mathbf{4.87 \pm 0.26}$

(b)

two agents and by 6% in case of three agents when compared with the parameters of two and three separate networks respectively. Furthermore, the addition of a single agent to our architecture reduces the required number of parameters by 6% compared to a single standalone agent. Due to the regularization effect that multiple agents have on their training and the implicit knowledge transfer, the training time our approach needs on average 25.000–50.000 less time steps to converge compared with a single DQN and each training epoch needs approximately 30 min less than the training of 2 epochs in a separate single DQN (NVIDIA Titan-X, 12 GB). Inference is on par with a single agent at ∼20fps.

4 Conclusion

In this paper we formulated the problem of multiple anatomical landmark detection as a multi-agent reinforcement learning scenario, we also introduced Collab-DQN, a Collaborative DQN for landmark detection in brain and cardiac MRI volumes and 3D US. We train K agents together looking for K landmarks. The agents share their convolutional layer weights. In this fashion we exploit the knowledge transferred by each agent to teach the other agents. We achieve significantly better performance than the next best method of [2] decreasing the error by more than 1 mm while taking less time to train and less memory than training K agents serially. We believe that a Bayesian exploration approach is a natural next step, which will be addressed in future work.

Acknowledgements. Wellcome Trust IEH Award [102431], EPSRC EP/ S013687/ 1, NVIDIA for their GPU donations. Brain MRI: adni.loni.usc.edu, US data: access only with informed consent, subject to approval and formal Data Sharing Agreement. Caridac data: digital-heart.org.

References

1. Alansary, A., et al.: Automatic view planning with multi-scale deep reinforcement learning agents. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 277–285. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_32
2. Alansary, A., et al.: Evaluating reinforcement learning agents for anatomical landmark detection. Med. Image Anal. **53**, 156–164 (2019)
3. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: Signature verification using a “siamese” time delay neural network, pp. 737–744 (1993)
4. Foerster, J., Assael, I.A., de Freitas, N., Whiteson, S.: Learning to communicate with deep multi-agent reinforcement learning. In: NIPS, vol. 29, pp. 2137–2145 (2016)
5. Foerster, J., Chen, R.Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., Mordatch, I.: Learning with opponent-learning awareness. In: Proceedings of 17th International Conference on Autonomous Agents and MultiAgent Systems AAMAS 2018, pp. 122–130 (2018)

6. Gauriau, R., Cuingnet, R., Lesage, D., Bloch, I.: Multi-organ localization with cascaded global-to-local regression and shape prior. Med. Image Anal. **23**(1), 70–83 (2015)
7. Ghesu, F., Georgescu, B., Zheng, Y., Grbic, S., Maier, A., Hornegger, J., Comaniciu, D.: Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. IEEE PAMI **41**(1), 176–189 (2019)
8. Ghesu, F.C., Georgescu, B., Mansi, T., Neumann, D., Hornegger, J., Comaniciu, D.: An artificial agent for anatomical landmark detection in medical images. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 229–237. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46726-9_27
9. Girard, J., Emami, R.: Concurrent Markov decision processes for robot team learning. EAAI **39**, 223–234 (2015)
10. Gupta, J.K., Egorov, M., Kochenderfer, M.: Cooperative multi-agent control using deep reinforcement learning. In: Sukthankar, G., Rodriguez-Aguilar, J.A. (eds.) AAMAS 2017. LNCS (LNAI), vol. 10642, pp. 66–83. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71682-4_5
11. Jaakkola, T., Singh, S.P., Jordan, M.I.: Reinforcement learning algorithm for partially observable Markov decision problems. In: NIPS (1995)
12. Jack Jr., C.R., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging **27**(4), 685–691 (2008)
13. Li, Y., et al.: Fast multiple landmark localisation using a patch-based iterative network. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 563–571. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_64
14. de Marvao, A., Dawes, T.J., Shi, W., Minas, C., Keenan, N.G., Diamond, T., Durighel, G., Montana, G., Rueckert, D., Cook, S.A., et al.: Population-based studies of myocardial hypertrophy: high resolution cardiovascular magnetic resonance atlases improve statistical power. J. Cardiovasc. Magn. Reson. **16**(1), 16 (2014)
15. Mnih, V., et al.: Human-level control through deep reinforcement learning. Nature **518**, 529 (2015)
16. Oktay, O., et al.: Stratified decision forests for accurate anatomical landmark localization in cardiac images. IEEE Trans. Med. Imaging **36**(1), 332–342 (2017)
17. Rahmatullah, B., Papageorghiou, A.T., Noble, J.A.: Image analysis using machine learning: anatomical landmarks detection in fetal ultrasound images. In: 2012 IEEE 36th Annual Computer Software and Applications Conference, pp. 354–355, July 2012
18. Rashid, T., Samvelyan, M., de Witt, C.S., Farquhar, G., Foerster, J.N., Whiteson, S.: QMIX: monotonic value function factorisation for deep multi-agent reinforcement learning. CoRR abs/1803.11485 (2018)
19. Zheng, Y., Liu, D., Georgescu, B., Nguyen, H., Comaniciu, D.: 3D deep learning for efficient and robust landmark detection in volumetric data. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 565–572. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_69