Reinforced Redetection of Landmark in Pre- and Post-operative Brain Scan Using Anatomical Guidance for Image Alignment

Diana Waldmannstetter[1,2(✉)], Fernando Navarro[1,4], Benedikt Wiestler[3], Jan S. Kirschke[3], Anjany Sekuboyina[1,3,4], Ester Molero[1], and Bjoern H. Menze[1,2,4]

[1] Department of Informatics, Technical University of Munich, Munich, Germany
diana.waldmannstetter@tum.de
[2] Munich School of BioEngineering, Munich, Germany
[3] Department of Neuroradiology, Klinikum rechts der Isar, Munich, Germany
[4] TranslaTUM - Central Institute for Translational Cancer Research, Munich, Germany

Abstract. Re-identifying locations of interest in pre- and post-operative images is a hard identification problem, as the anatomical landscape changes dramatically due to tumor resection and tissue displacement. Classical image registration techniques oftentimes fail in vicinity of the tumor, where the enclosing structures are massively altered from one scan to another. Still, locations nearby the tumor or the resection cavity are the most relevant for evaluating tumor progression patterns and for comparing pre- and post-operative radiomic signatures. We address this issue by exploring a Reinforcement Learning (RL) approach. An artificial agent is self-taught to find the optimal path towards a target driven by a feedback signal from the environment. Incorporating anatomical guidance, we restrict the agent's search space to surgery-unaffected structures only. By defining landmarks for each patient individually, we aim to obtain a patient-specific representation of its differential radiomic features across different time points for enhancing image alignment. Estimated landmarks reach a remarkable mean distance error around 3 mm. In addition, they show a high agreement with expert annotations on a challenging dataset of MR scans from the brain before and after tumor resection.

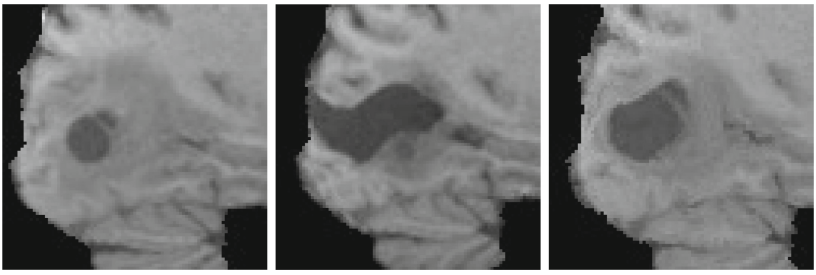
Keywords: Reinforcement Learning · Image registration · Image alignment · Differential radiomics · Brain tumor

1 Introduction

The most effective treatment for progression delay in aggressive primary brain tumors is tumor resection, usually followed by radiation therapy or chemotherapy [4]. When evaluating the post-operative scans, the areas that show signs of

Ž. Špiclin et al. (Eds.): WBIR 2020, LNCS 12120, pp. 81–90, 2020.
https://doi.org/10.1007/978-3-030-50120-4_8

tumor re-growth are compared to the same areas in the pre-operative scans. As there is almost always a shift in brain tissue as well as tumor- and resection-induced intensity changes, conventional image registration techniques oftentimes fail when it comes to map the differing structures, see Fig. 1. We aim to evaluate local patterns with anatomical guidance for a better adaption in this task. We perform re-identification of landmarks for making use of quantitative radiomic approaches, since radiomics intend to improve image analysis by extracting large amounts of quantitative features [8]. In order to detect reference points before and after tumor resection, we use individual landmarks for each patient, representing locations prone to progression. Redetecting these landmarks automatically in follow-up scans may simplify future image alignment for the same patient. Therefore, we define multiple patient-specific landmarks around the tumor and take a first step towards differential radiomic feature extraction and therefore, a more precise alignment of the resection-affected regions.



(a) pre-operative (b) post-operative (c) registered

Fig. 1. 2D zoom into tumor-/resection-affected regions in (a) a pre-operative scan, (b) the corresponding post-operative scan and (c) the result of a standard image registration

Recent literature shows a variety of approaches towards localization of anatomical landmarks in medical images. Li *et al.* [9] developed a patch-based CNN for landmark localization combining regression and classification for the detection of both single and multiple landmarks simultaneously by involving Principle Component Analysis (PCA). Zheng *et al.* [19] evolved a two-step approach combining a shallow network with a deep network for efficient landmark detection. Another two-stage approach was proposed by Zhang *et al.* [18] comprising a patch-based CNN regression model followed by another CNN for predicting landmarks in an end-to-end manner. For the first time, Ghesu *et al.* [6] introduced Deep RL for localizing anatomical landmarks using Q-learning. Their method is further developed by exploring multiple scales in [5,7]. In [11], Maicas *et al.* adopt this method and extend it to the more complex detection of breast lesions. There, adaptive bounding boxes are leveraged to train the agent. Alansary *et al.* [2] presented a multi-scale strategy by iteratively training their agent using action steps with different sizes on multiple scales. Additionally, they evaluate several Q-Learning approaches as there are Double, Dueling

and Dueling Double Q-Learning. A variant of their approach towards automatic view planning is shown in [1] and an extension towards the detection of multiple landmarks simultaneously in [15].

The recent success of RL in the field of landmark localization [2,5–7,15] in combination with the ability of RL agents to adapt to a specific environment, encouraged us to transfer this approach for landmark redetection to pre- and post-operative brain images. Furthermore, RL has the benefit of being able to perform on limited training data, which is crucial for our task. Based on the approach in [2], we further develop the method to consider anatomical guidance and propose the following contributions: First, we present two RL-based agents. A baseline agent and an extended version of it under anatomical guidance, improving the agent's ability to adapt to the issue of altering tissue structures by integrating patient-specific anatomy into our model. Second, we evaluate our approach on a challenging dataset of MR scans before and after tumor resection, provided by the BraTS challenge [12] and TCIA [3], achieving results comparable to an expert performance. Additionally, we provide annotations for this data.

In the following, we present how we utilize Q-Learning in RL (Sect. 2) and introduce our extension (Sect. 3), before demonstrating the performance of our approach on a complex data set.

2 Deep Reinforcement Learning Using Q-Learning

In RL, an artificial agent is self-taught by interacting with an environment. In every step, the agent retrieves a reward from its environment after executing an action. The final goal of the agent is to find an optimal policy, guiding the agent from any given state to the target by maximizing future rewards. This can be formulated as a sequential decision process. RL then is modeled as a *Markov Decision Process* (MDP), which defines the interaction between the agent and its environment. The agent executes an action $a \in A$ at state $s \in S$, returning a reward signal $r \in R$ at each time step t [14].

Finding the optimal policy is described by the action-value function $Q(s, a)$, which is optimized during training and gives the maximum expected discounted future reward, where the accumulated discounted reward after τ time steps is defined as

$$R_\tau = \sum_{\tau=0}^{\infty} \gamma^\tau r_{t+\tau+1}, \tag{1}$$

with the discount rate $\gamma \in [0, 1]$ for weighting immediate and future rewards [17]. Using the Bellman optimality equation, the action-value function can be solved recursively [14]:

$$Q(s, a) = \mathbb{E}\left[r + \gamma \max_{a'} Q(s', a')\right], \tag{2}$$

where s' and a' are the possible subsequent state and action.

Mnih *et al.* [13] developed the Deep Q-Network (DQN), which approximates

$$Q(s,a) \approx Q(s,a;\theta), \tag{3}$$

using a CNN with the network parameters θ. For stability reasons, a target network $Q(\theta^-)$ is introduced. It estimates the actual Q-network iteratively by updating the parameters of the target network only every nth iteration with the steadily updated Q-network parameters. The loss function reads:

$$L_n(\theta_n) = E_{s,a,r,s'}\left[\left(r + \gamma \max_{a'} Q(s',a';\theta_n^-) - Q(s,a;\theta_n)\right)^2\right] \tag{4}$$

Experience replay technique [10] is added, training the network using randomly sampled minibatches from experiences the agent has already gained. This is stored in an experience replay memory.

The DQN was further improved by Wang *et al.* [16], separating the network into two partitions. One handles the state-value function $V(s)$ and the other one deals with the advantage function $A(s,a)$, see Fig. 2. Both are then combined by an aggregation layer to provide a single Q-function

$$Q(s,a) = V(s) + A(s,a). \tag{5}$$

Here, estimating the state-value function is essential in every time step, while this is not necessary for the advantage function. Consequently, the dueling network learns the state-value function more accurately, thus improving the network performance with increasing number of actions.

3 Anatomically Guided RL Agent

Similar to [2], we make use of a Deep Q-Network with dueling architecture, shown in Fig. 2. Each state in our image environment is modeled as a 3D patch centered around the current location of the agent, see Fig. 3(a). Hence, the agent sees a different part of its environment in every time step. Due to the experience replay technique, we define an experience buffer storing the last four patches, which the network can see in one iteration, enhancing the agent's robustness.
We define the action space with the six actions *right, left, forward, backward, up, down*. This results in two actions along each axis in positive and negative direction, $a \in A = \{+x, -x, +y, -y, +z, -z\}$, thus moving the agent by one voxel. The reward is defined similar to [6], calculating the relative change in the distance to the position of the target landmark. Furthermore, we make use of a search strategy operating on multiple scales [2,5,7] for more robustness and efficiency.

A key feature of our approach is the anatomical guidance. We return a negative reward $r = -1$ when the agent steps inside the surgery-affected regions. Therefore, we provide the segmentation mask to the agent, so the agent learns to stay in unaffected structures only, see Fig. 3(b). Since this guides the agent to move towards the target without touching the immense tissue changes inside the most affected regions, this leads to a policy that is more generalizable to altering brain tissue.

4 Experimental Setup

We evaluate our approach on a challenging dataset provided by the BraTS challenge [12] and TCIA [3]. We use MR image data from 10 patients with brain tumors, with one scan before and one scan after tumor resection, comprising 20 image volumes in total. All images are skull-stripped, rigidly co-registered and interpolated to a common resolution of 1 mm^3, while the initial resolution is in the range of 3–8 mm for most sequences. The dataset includes the image volumes and their corresponding segmentation masks of the tumor- and resection-affected regions in the pre- and post-operative scans, respectively. For each patient, 3 landmarks in the post-operative scan are annotated by a clinical expert, in varying distances up to 4 cm around the resection-affected region. The same expert redetected the landmarks in the corresponding pre-operative scan for generating ground truth annotations.

Training and Testing. Before training, we crop an initialization box of size $50 \times 50 \times 50$ voxels around the target in the training image, see Fig. 3(a). When training under anatomical guidance, we exclude the resection mask, see Fig. 3(b). Then, we randomly initialize the agent inside this region and sample a patch of size $15 \times 15 \times 15$ voxels, which follows the agent in every step. For every patient and landmark individually, we train on the respective post-operative scan and test on the corresponding pre-operative scan, generating patient-specific models. Similar to [2], we define the terminal state in training as the point, when the distance between the agent and the target landmark is less or equal to 1 mm. During testing, the agent is stopped, when it is oscillating around the same location.

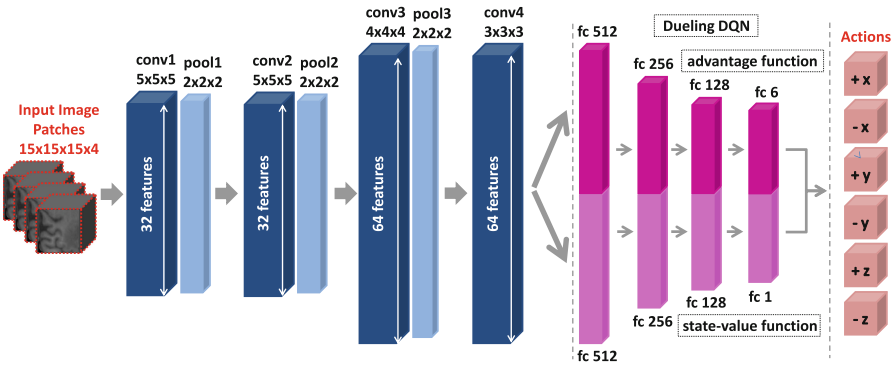


Fig. 2. Our network architecture with a dueling DQN. A 3D patch is sampled around the current position of the agent and fed to the network, consisting of convolutional (conv) layers alternating with pooling (pool) layers, followed by a dueling DQN with fully connected (fc) layers. The network outputs the Q-value for the six possible actions, whereof the agent selects the one with the highest value.

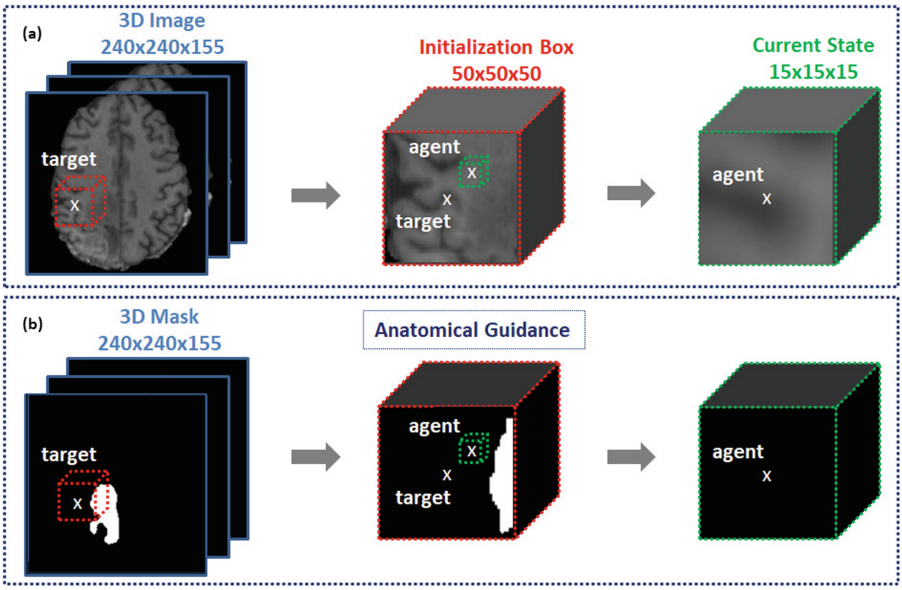
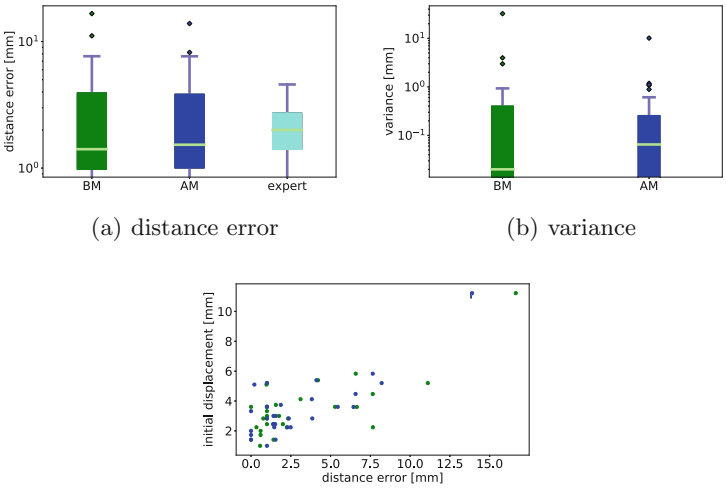


Fig. 3. (a) Patch generation. A box of size $50 \times 50 \times 50$ is sampled around the target for initialization reasons. Then, a patch of size $15 \times 15 \times 15$ is extracted around the current position of the agent, representing the current state. (b) Anatomical Guidance. We provide the segmentation mask to the agent for excluding this region during training. When stepping inside the masked region, a negative reward is returned, so that the agent is guided to avoiding affected structures.

Experiments. We use two different agents for each experiment. One is trained on the baseline method without anatomical guidance, whereas the other one is anatomically guided. Due to the lack of a validation set, we tune the model on the respective training image. For each experiment, we then choose the model that is performing best on the corresponding test image. Although this might lead to some underestimation regarding the distance error measurements, it makes sure that we provide the same conditions for the two agents in the different experiments, leading to comparable results. During evaluation, we define 20 fixed starting points, typically converging to slightly different final endpoints. Since we initialize the training agent inside the initialization box, we use the same box for testing and select the starting points from there. For each of the 20 evaluation runs per experiment, we calculate the Euclidean distance in mm between the final location of the agent and the true landmark, which gives us the distance error, and calculate the mean, for producing comparable results. Subsequently, for the sake of simplicity, we refer to this mean distance error simply as distance error.

5 Results

Quantitative results can be observed from Fig. 4(a) and (b), showing the distance errors in mm for the baseline method (BM), the extended method using anatomical guidance (AM), as well as another expert's annotation for comparison. Therefore, we take the landmark annotations of a second expert, when performing the redetection task manually, and calculate the Euclidean distance to the ground truth annotations, giving us the distance errors of an expert. For further comparisons, we calculate additional measurements on the distance errors, the mean and the median distance as well as the normalized mean and median, respectively, see Table 1. All measurements are calculated on landmark level. Qualitative results are presented in Fig. 5, showing a sample redetection for two different landmarks, where both methods achieve high precision with a distance error of 0 mm.



(a) distance error (b) variance

(c) initial displacement

Fig. 4. Results for the baseline method (BM, green), the extended method with anatomical guidance (AM, dark blue) and an expert annotation for comparison (light blue). (a) shows the distance mean errors in mm and (b) the variances in the distance errors due to the multiple starting points. (c) shows the relation between the initial expert annotation displacements and the distance errors. The dots represent the respective offsets of the initial expert annotation for the training and test images in relation to the corresponding distance errors of BM and AM. (Color figure online)

RL vs Expert. The lowest distance errors for both methods are close to 0 mm, representing a perfect redetection. The highest lie above 1 cm. Due to the 20 starting points, we achieve variances in the distance errors, tending to increase with growing errors, see Fig. 4(a) and (b). High variances are caused by some

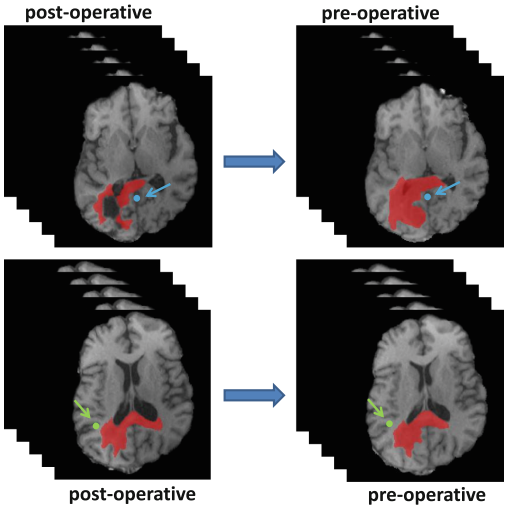


Fig. 5. Sample redetection for two different landmarks (top, bottom) in the same patient. The masked region is marked in red. The landmarks in blue and green, respectively. (Color figure online)

outliers, where the agent gets lost in the environment. However, it is remarkable that we achieve a variance of 0 in some experiments, which means that the agent navigates towards the same target from every starting point, demonstrating high robustness. As Fig. 4(c) shows, the distance errors from BM and AM both scale with the initial displacements between the ground truth annotations in the training and test images, when annotated by an expert. That means, a larger offset between the initial expert landmark annotation in the training and the test image results in larger distance errors. This makes sense, since larger initial annotation displacements are linked to larger tissue changes. Nevertheless, the majority ranges within smaller errors from 0–4 mm. From Table 1, we observe that the mean of all distance errors is lowest for the comparison expert, while both BM and AM show high agreement with it. Still, the median of all distance errors is smaller for BM and AM. A normalization with the initial displacements leads to similar mean errors of both RL methods and the comparison expert, see Table 1.

Benefits of Anatomical Guidance. Figure 4(a) and (b) as well as Table 1 show that AM performs more robust than BM, since the outliers have slightly smaller mean distance errors and variances. Hence, incorporating anatomical guidance outperforms the baseline agent in average, while showing high agreement with the comparison expert's annotations. Our approach achieves noticeable performance with an average distance error below 3 mm. Moreover, anatomical guidance provides potential to incorporate additional anatomical information.

Table 1. Calculations on the distance errors for BM, AM and an expert annotation.

	Mean distance [mm]	Median distance [mm]	Normalized mean	Normalized median
BM	3.05	1.41	0.79	0.57
AM	2.82	1.53	0.74	0.64
Expert	2.18	2.0	0.75	0.72

6 Conclusion

In this work, we presented a RL framework for landmark redetection in a challenging dataset of pre- and post-operative brain scans. We evaluated two RL agents: a basic one exploring the full environment and an extended one guided by the resection anatomy for finding the optimal path towards the target landmark. Overall, both approaches showed good results in terms of speed and accuracy, while the agent under anatomical guidance performs better in average. Therefore, this approach allows to further develop the guidance by anatomical structures, especially in analyzing the connection between different time points before and after tumor resection, for generating a more representative and efficient model of the anatomical changes. For further automatization, the segmentation masks can be produced using some segmentation framework, which would be of minimal additional effort here and would be needed to be done once for training only. Additionally, we will invest in finetuning our approach towards a more robust redetection for eliminating outliers. Moreover, we will further investigate in generating a dense representation of patient-specific differential radiomics by localizing multiple landmarks simultaneously, ideally incorporating the spatial relationships between tumor structures, resection region and landmarks.

Acknowledgements. Supported by Deutsche Forschungsgemeinschaft (DFG) through TUM International Graduate School of Science and Engineering (IGSSE), GSC 81.

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., Oktay, O., Li, Y., et al.: Evaluating reinforcement learning agents for anatomical landmark detection. Med. Image Anal. **53**, 156–164 (2019)
3. Clark, K., Vendt, B., Smith, K., et al.: The cancer imaging archive (TCIA): maintaining and operating a public information repository. J. Digit. Imag. **26**(6), 1045–1057 (2013)
4. DeAngelis, L.: Brain tumors. New Engl. J. Med. **344**(2), 114–123 (2001)

5. Ghesu, F.C., Georgescu, B., Grbic, S., Maier, A.K., Hornegger, J., Comaniciu, D.: Robust multi-scale anatomical landmark detection in incomplete 3D-CT data. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 194–202. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_23
6. 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
7. Ghesu, F., et al.: Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. IEEE TPAMI **41**(1), 176–189 (2017)
8. Lambin, P., Rios-Velazquez, E., Leijenaar, R., et al.: Radiomics: extracting more information from medical images using advanced feature analysis. Eur. J. Cancer **48**(4), 441–446 (2012)
9. 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
10. Lin, L.J.: Self-improving reactive agents based on reinforcement learning, planning and teaching. Mach. Learn. **8**(3–4), 293–321 (1992)
11. Maicas, G., Carneiro, G., Bradley, A.P., Nascimento, J.C., Reid, I.: Deep reinforcement learning for active breast lesion detection from DCE-MRI. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 665–673. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_76
12. Menze, B., Jakab, A., Bauer, S., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imag. **34**(10), 1993–2024 (2015)
13. Mnih, V., Kavukcuoglu, K., Silver, D., et al.: Human-level control through deep reinforcement learning. Nature **518**(7540), 529 (2015)
14. Sutton, R., Barto, A.: Introduction to Reinforcement Learning, 1st edn. MIT Press, Cambridge (1998)
15. Vlontzos, A., Alansary, A., Kamnitsas, K., Rueckert, D., Kainz, B.: Multiple landmark detection using multi-agent reinforcement learning. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 262–270. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_29
16. Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., De Freitas, N.: Dueling network architectures for deep reinforcement learning. In: International Conference on Machine Learning, pp. 1995–2003 (2016)
17. Watkins, C., Dayan, P.: Q-learning. Mach. Learn. **8**(3–4), 279–292 (1992)
18. Zhang, J., Liu, M., Shen, D.: Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. IEEE Trans. Image Process. **26**(10), 4753–4764 (2017)
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