International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 10, ISSN No. 2455-2143, Pages 139-149
Published Online February 2024 in IJEAST (http://www.ijeast.com)

REINFORCEMENT LEARNING IN MEDICAL IMAGE ANALYSIS

Dr. T. Amitha
Professor /CSE Dept,
SVS GROUPS OF INSTITUTION,
Warangal, Telangana

Abstract: **Motivation: Medical image analysis involves tasks to assist physicians in qualitative and quantitative analysis of lesions or anatomical structures, significantly improving the accuracy and reliability of diagnosis and prognosis. Traditionally, these tasks are finished by physicians or medical physicists and lead to two major problems: (i) low efficiency; (ii) biased by personal experience. In the past decade, many machine learning methods have been applied to accelerate and automate the image analysis process. Compared to the enormous deployments of supervised and unsupervised learning models, attempts to use reinforcement learning in medical image analysis are scarce. This review article could serve as the stepping-stone for related research.**

Significance: From our observation, though reinforcement learning has gradually gained momentum in recent years, many researchers in the medical analysis field find it hard to understand and deploy in clinics.One cause is lacking well-organized review articles targeting readers lacking professional computer science backgrounds. Rather than providing a comprehensive list of all reinforcement learning models in medical image analysis, this paper may help the readers to learn how to formulate and solve their medical image analysis research as reinforcement learning problems.

I. INTRODUCTION

The purpose of medical image analysis is to mine and analyze valuable information from medical images by using digital image processing to assist doctors in making more accurate and reliable diagnoses and prognoses. According to different imaging principles, common imaging modalities can be categorized as CT, MR, Ultrasound, SPECT, PET, X-ray, OCT, and microscope. Medical image processing can also be classified according to specific processing tasks. Typical tasks include classification, segmentation, registration, and recognition. Figure 1 shows the range of our review article.

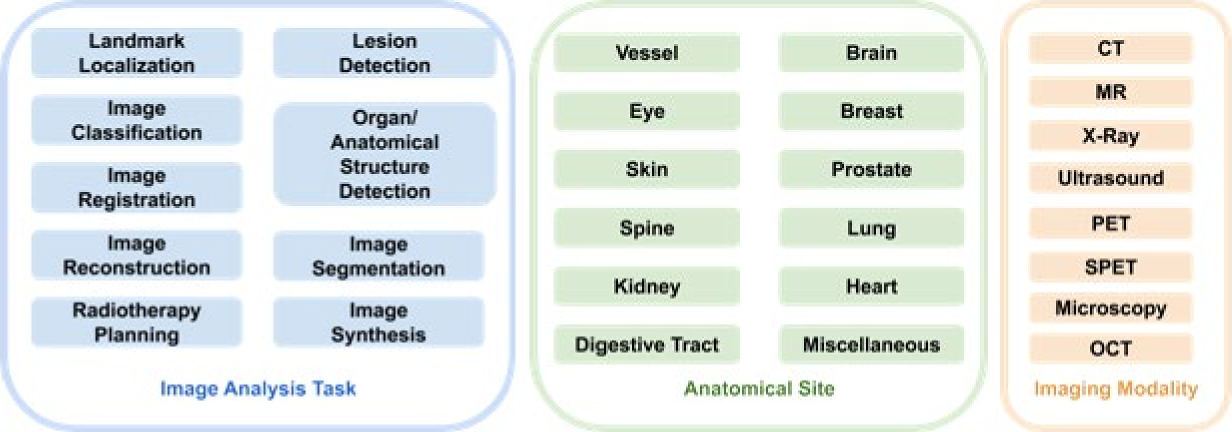

Figure 1: Range of our review article. Blue box: covered image analysis tasks; green box: covered anatomicalsites; yellow box: covered imaging modalities.

With the development of imaging technology and the iterative update of imaging equipment, the time required for medical imaging is greatly shortened, and the resolution of imaging is also significantly improved. At the same time, the data volume of medical images has experienced an unprecedented surge, with the trend of high dimensionality. The traditional manual analysis of medical images by physicians became tedious and inefficient. More and more physicians are looking to automate this process by partnering with engineers. That’s how the combined medical imaging and machine learning field was born.Many excellent algorithms in the field of natural image analysis have also shown good results in the field of medical images (Shen et al., 2017).

Reinforcement learning (RL) is neither supervised learning nor unsupervised learning. The goal of reinforcement learning is to achieve the maximum expected cumulative reward (Sutton & Barto, 2018).

Figure 2 shows the relationship between machine learning, supervised learning, unsupervised learning, reinforcement learning and deep learning.

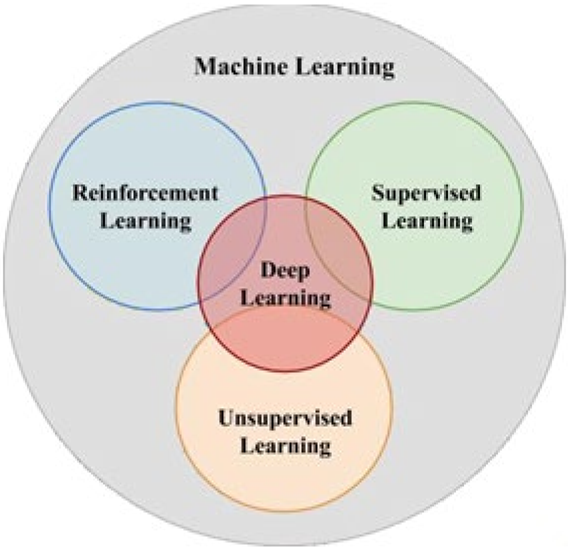

Figure 2: Relationship between machine learning, supervised learning, unsupervised learning andreinforcement learning.

The number of published reinforcement learning-related papers has grown rapidly in the past two decades. State-of-the-art RL models have been applied to solve problems that are difficult or infeasible with other machine learning approaches, such as playing video games (Mnih et al., 2013; Mnih et al., 2015; Silver et al., 2017), natural language processing (Sharma & Kaushik), and autonomous driving (Sallab et al., 2017). These RL methods have achieved outstanding performances. However, attempts to exploit the technical developments in RL in the medical analysis field are scarce. Figure 3 shows the trends of number of published machine learning papers and reinforcement learning papers in medical image analysis. Despite the overall growth trend, the number of published RL papers still only constitutes a tiny part of machine learning in medical image analysis. On the other hand, RL methods have unique advantages in dealing with medical image data:

RL models can efficiently learn from limited annotation guided by supervisedactions step by step, while medical data often lacks large-scale accessible annotation.

RL models are less biased since they won't inherit bias from the labels made byhuman annotators.

RL-agents can learn from sequential data, and the learning process is goal- oriented. Besides exploiting experience, it can also explore new solutions. The RL can even surpass human experts when solving the same problem.

The review article is based on Synthesis Methodology (Wilson & Anagnostopoulos, 2021).

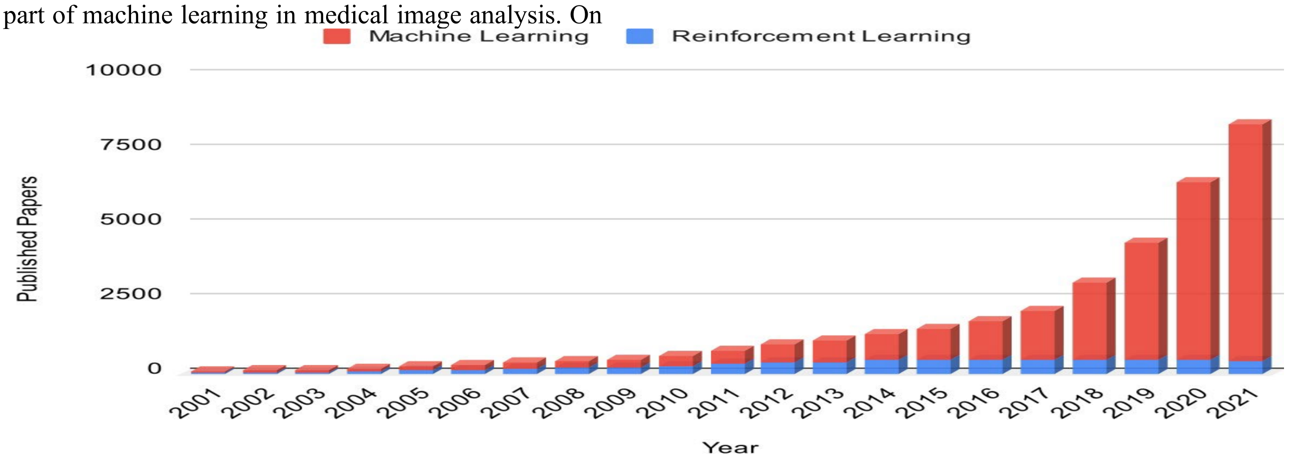

Figure 3: Trends of number of published machine learning papers and reinforcement learning papers in medical image analysis.

This figure is made by separately searching the keywords "Machine Learning AND (Medical Imaging OR (Medical Image Analysis))" and "Reinforcement learning AND (Medical Imaging OR (Medical Image Analysis))" in PubMed. The number of papers published each year is counted.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) will be followed(Moher et al., 2009). Firstly, the following pattern will be searched in Google Scholar and PubMed:Clustering AND (Medical OR CT OR MR OR Ultrasound OR X-ray OR OCT) AND IMAGE AND Segmentation. Then the duplicate papers will be removed. We set the qualified publication date to 2010. The remaining papers will go through qualitative synthesis and quantitative synthesis. The summary of the selection process is shown in Figure 4.

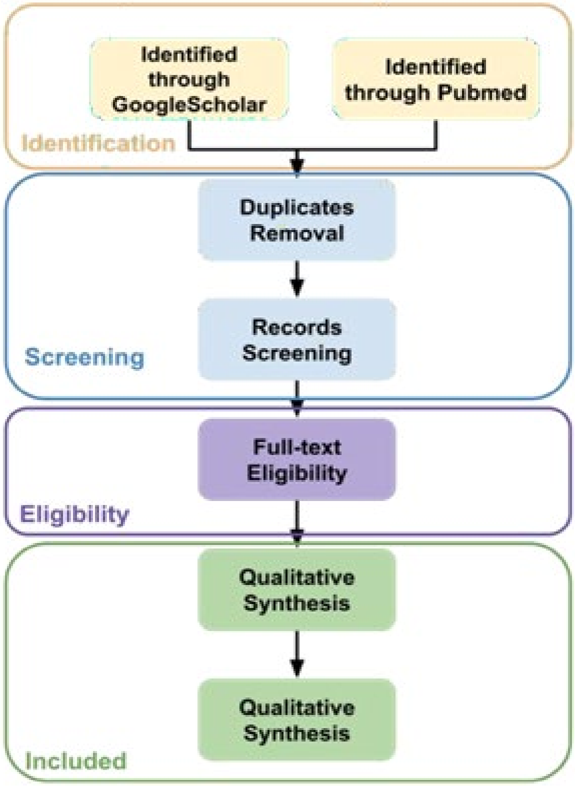

Figure 4: Flows of information through the different phases of a systematic review.

By reviewing content, analyzing common points and comparing difference of these papers, we hope that we can inspire our target readers to (i) have a better understanding of RF, (ii) learn how to formulate their research problems as RL problems. For the next two sections, we will first prepare the readers with basic knowledge of RL. Then we will show how to apply RL in different medical image analysis tasks. Those readers who have already been familiar with RL algorithms could directly go to the application section.

1. Reinforcement Learning Basics

In this subsection, we provide a list of terminologies that frequently appear in RL papers. Some terminologies may appear in definitions of other terminologies before they are defined.

- Action (A): An action (a) is the way that an agent interacts with the environment. A includes all possible actions that an agent could perform.
- Agent: Agents are the models we attempt to build that interact with the environment and takeactions.
- Environment: The content that the agent is interacting with is called the environment. Whileproviding feedback after the agent takes action, the environment itself is also changing.
- State (S): A state (s) is a frame of an environment. S includes all states that an agent will gothrough.
- Reward (R): A positive reward (r) means an increase possibility of achieving the goal, whilea negative reward means the decreased possibility. R includes all the possible reward values the environment may feed back to the agent.
- Episode: If an agent has gone through all the states from the initial state to the terminal state,we say this agent has finished the episode.
- Transition probability $P(s'|s, a)$: P(s'|s, a) is the

possibility of transiting to transiting tostate s' to from the current state s, taking the action a.

- Policy $\pi(a|s)$: The policy instructs the agent to choose among actions A under the currentstate.
- Return (G): The return is the cumulative discounted future reward.
- $G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2}$, where t is the time and γ is the discount factor.
- State value $V^\pi(s)$: The expected amount of return from current state.
- $V^\pi(s) = E[G_t|s_t = s]$, where E is the expectation.
- Action value $Q^\pi(s, a)$ (Q value): The expected amount of return from current state, taking action s. $Q^\pi(s, a) = E[G_t|s_t = s, a_t = a]$
- Optimal action value: $Q^\star(s, a)$: $Q^\star(s, a) = max\ Q^\pi(s_t, a_t)$

π

- Agent environment interaction: Figure 5 shows how the agent is interaction with theenvironment.

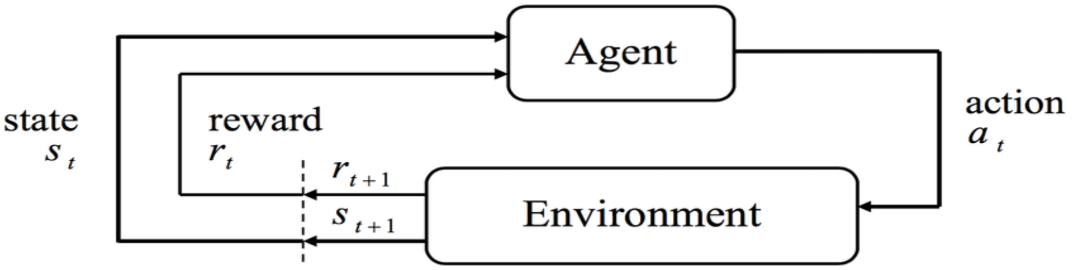

Figure 5: Agent Environment Interaction. Adapted from (Rafati & Noelle, 2019).

With the development of the RL theory, numerous algorithms have been created. Benefiting from the combination with deep learning, RL is now capable of handling more and more complex scenarios in modern applications. But no matter how complex these state-of-the-art algorithms are, they can be mainly divided into two categories: model- based RL and model-free RL. As its name indicates, model-based RL attempts to explain the environment and create a model to simulate it. Model-free RL, however, will only update its policy by interacting with the environment and observing the rewards.

We can further divide the model-free RLs into policy-based and value-based according to whether the algorithm is optimizing the value function or policy. Value-based RLs are widely applied for discrete action space problems, while policy-based RLs are suitable for both discrete and continuous action space. Some RL algorithms are based on both the value and policy, like DDPG (Lillicrap et al., 2015), TD3 (Fujimoto etal., 2018) and SAC (Haarnoja et al., 2018). Figure 6 shows the taxonomy of popular RL algorithms. In ourreview, all the RL models are model-free, and the mostly used algorithms are DQN, DDQN, A2C, and DDPG. Below we include brief introductions of these RL algorithms commonly used in medical image analysis.

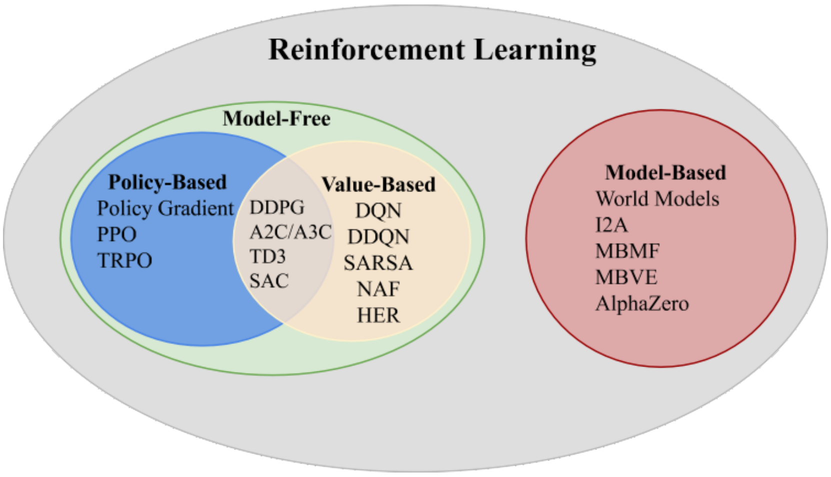

Figure 6: Reinforcement learning algorithms taxonomy.

DQN

The Deep Q-Network (DQN) was first proposed by (Mnih et al., 2013; Mnih et al., 2015) to solve somecomplex computer perception vision problems. It combined the idea of the traditional Q learning method (Watkins, 1989) and the deep CNN (Krizhevsky et al.). The motivation of DQN is to solve the problem that the Q-table can only store a limited number of states, while in real-life scenarios, there could be an immense or even infinite number of states. DQN adopts the experience replay mechanism that randomly samples a small batch of tuples from the replay buffers during the training process. The correlations between the samples are significantly reduced, leading to better algorithm robustness. Another improvement, compared to Q learning, is that DQN uses a deep CNN to represent the current Q function and uses another network to define the target Q value. The introduction of the target Q value network reduced the correlation between the current and target Q values. Figure 7 shows the workflow of the DQN.

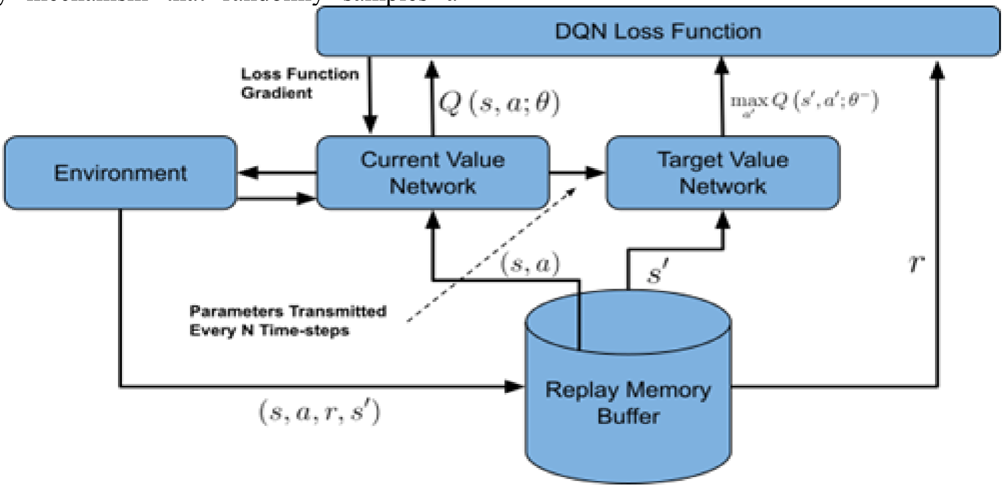

Figure 7: Workflow of DQN algorithm.

DDQN

DQN is one of the most popular RL algorithms applied in medical image analysis. How- ever, theoptimization target in DQN is represented as $r + \gamma \max Q(s', a'|\theta^-)$. The selection and evaluation of actions are all based on the network's same parameter, leading to over- estimation of the Q value. The Double DQN (DDQN), which (Van Hasselt et al.) first proposed, used two separate networks for selection and evaluation. Here the target Q value is written as $r + \gamma Q(s'$, argmax better more stable learned policy than DQN.

$Q(s', a|\theta_i), \theta^-)$, which achieved

2. RL in Medical Image Analysis

Medical Image Detection

Anatomical landmarks are biological coordinates that can be reallocated repeatedly and precisely on images produced by different imaging modalities — computed tomography (CT), ultrasound (US), magnetic resonance imaging (MRI). The accurate detection of anatomical landmarks is the ground for further medical image analysis tasks. Figure 8 is an example of vocal tract landmarks from the MRI image.

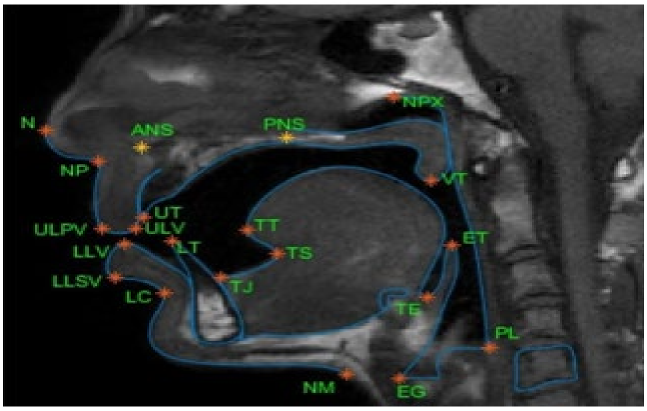

Figure 8: Vocal tract landmarks from MRI image Courtesy of (Eslami et al., 2020).

Many automatic algorithms for anatomical landmark detection have existed long before the attempts of using RL models. However, landmark detection, especially 3D landmarks detection, could be challenging and cause the failure of these algorithms (Ghesu et al., 2019). Moreover, the computation of features and hyper-parameters selection of the system may not be optimal since the involvement of human decisions. The researchers attempted a different paradigm to address this problem - translate the landmark detection tasks as reinforcement learning problems which is the common goal of the papers we reviewed. While most essential and tricky task in these papers, as you can see later, is designing the state space, action space, and reward space before training the models.

(Ghesu et al., 2016) is one of the very first papers that attempted to use RL for anatomical landmark detection. In an image I, $\vec{p}_{GT}$ denotes the location of anatomical landmark, and $\vec{p}_t$ denotes the location at the current time. State space S is the collection of all possible states $s_t = I(\vec{p}_t)$. Action space A is the collection of all possible actions by which the agent can move to the adjacent position, as illustrated by

Figure 9. Reward space R is defined as $\|\vec{p}_t - \vec{p}_{GT}\|2 - \|\vec{p}_{t+1} - \vec{p}_{GT}\|$ which impels the agent to move closer to the target anatomical landmark. A deep learning model was applied to approximate the state value function. The parameters are updated according to gradient descent, and the error function is:

θ= arg min E

$F\ [(y - Q(s, a; \theta\ ))\ \ ] + E$
$[V\ F\,(y)]$ (1)

$s,a,r,si\ \ s,a,r\qquad s$

This deep Q learning-based method beat the existing top systems not only in accuracy but also in speed. Thedesign of action, state, and reward spaces in the paper we just discussed became a standard method.

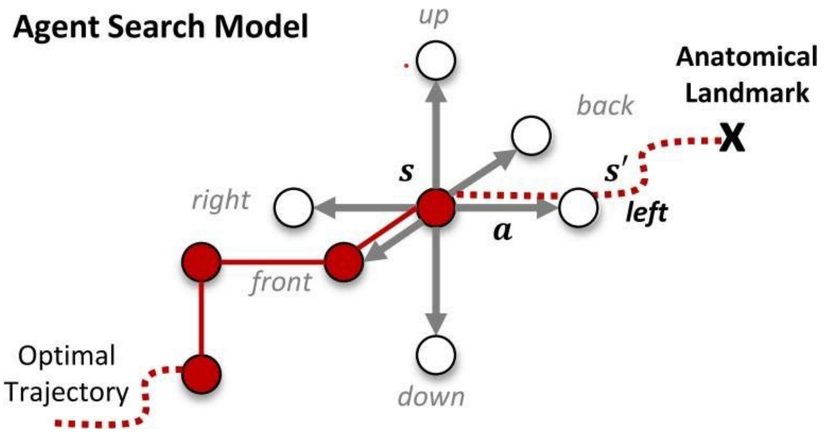

Figure 9: Possible actions of a 3D landmarks detection task. Courtesy of (Ghesu et al., 2019).

However, the approach mentioned above is still preliminary. One of the biggest disadvantages is thatit could not fully use the information at different scale. So a multi-scale deep reinforcement learning methodwas soon proposed in (Ghesu et al., 2019). The search for the landmark started from the coarsest scale. Once the search is convergent, the continued work would be started at a finer scale until the search meets the finest scale's convergence criteria. Figure 10 illustrates this non-trivial search process. Where L_d is the scale level in the continuous scale-space L, which can be calculated as:

$$L_d(t) = \psi_\rho(\sigma(t-1) * L_d(t-1)) \qquad (2)$$

Where ψ_ρ is the signal operator, and σ is the Gaussian-like smoothing function.

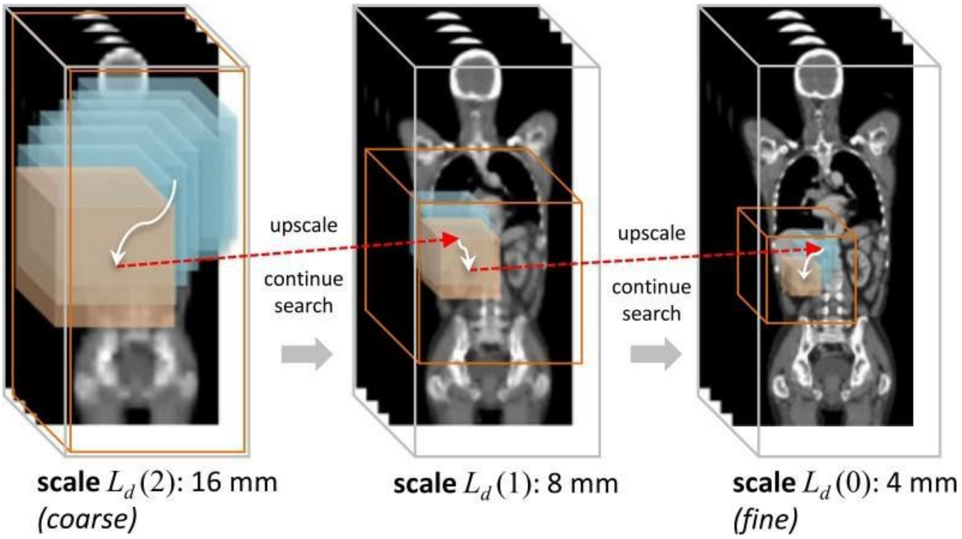

Figure 10: The trajectory of search the anatomical landmark across images of multiple scale-levels. Courtesyof (Ghesu et al., 2019).

(Alansary et al., 2019) extended the work of Ghesu et al. by evaluating different types of RL agents. He compared the detection results of using DQN, double DQN (DDQN), Double DQN, and duel double DQN (duel DDQN) on three different-modalities dataset — fetal US, cardiac MRI, and brain MRI.

Rather than detecting a landmark per agent separately, bold attempts have been made by (Vlontzos et al., 2019) to detect multiple landmarks with multiple collaboration agents. With the assumption that the anatomical landmarks have inner correlations with each other, the detection of one landmark could indicatethe location of some other landmarks. For the action function approximator in this paper, the collaborative deep Q network (Collab-DQN) was proposed. The weights of the convolutional layers are shared by all theagents, while the fully connected layers for deciding the actions are trained separately per agent. Compared to the methods that trained agents for different landmarks differently, this multi-agent approach reduced 50% detection error using a shorter training time.

Some other contributions to the RL for anatomical landmarks detection include: estimating the uncertainty of reinforcement learning agent (Browning et al., 2021), reducing the needed time to reach the landmark by using a continuous action space (Kasseroller et al., 2021), localization of modality invariant landmark (Winkel et al., 2020).

Lesion Detection

Object detection, also called object extraction, is the process of finding out the class labels and locationsof target objects in images or videos. It is one of the primary tasks in medical image analysis (Li et al., 2019). An exemplary detection result can be used as the basis to improve the performance of further taskslike segmentation.

The mainstream approaches for lesion detection nowadays still rely on exhaustive search methods thatcost a lot of time and deep learning methods that require a large amount of labeled data. Facing the current challenges and inspired by similar problems in landmarks detection (Ghesu et al., 2016), (Maicas et al., 2017) implemented a deep Q-network (DQN) agent for active breast lesion detection. The states are defined as current bounding box volumes of the 3D DCE-MR images. The reinforcement learning agent could gradually learn the policy to choose among actions to transit, scale the bounding box, and finally localize the breast lesion. Specifically, the action set consists of 9 actions that can translate the bounding box forward or backward along the x, y, z-axis, scale up or scale down the bounding box, and trigger the terminal state. To further evaluate the effectiveness of applying reinforcement learning on lesion detectionwith limited data, using DQN as the agent to localize brain tumors with very small training data was attempted by (Stember & Shalu, 2020), (Stember & Shalu, 2021a).Different from (Maicas et al., 2017), thebrain MR data are 2D slices. The environment is defined as the 2D slices overlaid with gaze plots viewed by the radiologist. Instead of using the bounding box, the states are the gaze plots the agent located. Threeactions - moving anterograde, not moving, moving retrograde would help the agent transfer to the next state. If the agent moves toward the lesion, it will receive a positive reward, otherwise a negative one. If the agent stays still, it will receive a relatively large positive reward within the lesion area or a rather large penalization otherwise. The experiment results showed that reinforcement learning models could work as robust lesion detectors with limited training data, reduce time consumption, and provide some interpretability.

Also addressing the lack of labeled training data, (Pesce et al., 2019) exploited visual attention mechanisms to learn from a combination of weakly labeled images (only class label) and a limited numberof fully annotated X-ray images. This paper proposed convolutional networks with attention feedback (CONAF) architecture and a recurrent attention model with annotation feedback (RAMAF) architecture. The RAMAF model can only observe one part of the image, which is defined as a state at a glimpse. The reinforcement learning agent needs to learn the policy to take a sequence of glimpses and finally locate thelesion site within the shortest time. Each glimpse consists of two image patches sharing the same central point, and the length of the glimpse sequence is fixed to be 7. The rewards will be decided according to (i)if the image is classified correctly; (ii) if the central point of a glimpse is within the labeled bounding box. RAMAF achieved a localization performance of detecting 82% of overall bounding boxes with a much faster detection speed than other state-of-the-art methods.

More than detecting lesions in static medical images (2D or 3D), the reinforcement- learning-based system can also track the lesions frame by frame continuously. (Luo et al., 2019) proposed a robust RL- based framework to detect and track plaque in Intravascular Optical Coherence Tomography (IVOCT) images. Despite the pollution problem of speckle- noise, blurred plaque edges, and diverse intravascular morphology, the proposed method achieved accurate tracking and has strong expansibility.

Three different modules are included in the proposed framework. The features are extracted and encoded first by the encoding feature module. Then the information of scale and location of the lesion is provided by the localization and identification module. Another function of this module is preventing over-tracking. The most important module is the spatial- temporal correlation RL module. Nine different actionsare different, including eight transformation actions and one stop action. The state S is defined as three- tuples: $S = (E, HL, HA)$. Here, the E represents the encoded output features from the FC1 layer. HL is thecollection of recent locations and scales. HA represents the recent ten sets of actions. 8000 IVOCT images were used to evaluate the framework. With a strict standard (IOU > 0.9), the RL module could improve theperformance of plaque tracking both frame-level and plaque-level.

Organ/ Anatomical Structure Detection

Besides detecting lesions, reinforcement learning can also be applied in organ detection. (Navarro et al., 2020) designed a deep Q-learning agent to locate various organs in 3D CT scans. The state is defined as voxel values within the current 3D bounding box. Eleven actions, including six translation actions, two zooming actions, and three scaling actions, make sure that the bounding box can move to any part of the 3D scan. The agent is rewarded if an action improves the intersection over union score (IOU). Seventy scans were used for training, and 20 scans were used for testing on seven different organs: pancreas, spleen,liver, lung (left and right), and kidney (left and right). This proposed method achieved a much faster speed than the region proposal and the exhaustive search methods and led to an overall IOU score of 0.63.

(Zhang et al., 2021) managed to detect and segment the vertebral body (VB) simultaneously. The sequence correlation of the VB is learned by a soft actor-critic (SAC) RL agent to reduce the background interference. The proposed framework consists of three modules: Sequential Conditional Reinforcement Learning network (SCRL), FC-ResNet, and Y-net. The SCRL learns the correlation and gives the attentionregion. The FC-ResNet extracts the low-level and high-level features to determine a more precise boundingbox according to the attention region. At the same time, the segmentation result is provided by the Y-net. The state of the RL agent is determined by a combination of the image patch, feature map, and region mask.And the reward is designed according to the change of attention-focusing accuracy to elicit the agent to achieve a better detection performance. This proposed approach accomplished an average of 92.3% IOU on VB detection and an average of 91.4% Dice on VB segmentation.

The research of (Zheng et al., 2021) was the first attempt to use the multi-agent RL in prostate detection. Two DQN agents locate the lower-left and upper-right corners of the bounding box while sharing knowledge according to the communication protocol (Foerster et al., 2016). The final location of the prostate is searched with a coarse-to-fine strategy to speed up the search process and improve the detectionaccuracy. In more detail, the agents first search on the coarsest scale to draw a big bounding box and gradually move to a finer scale to generate a smaller and more accurate bounding box to detect the prostate.Compared to the single-agent strategy (63.15%), this multi-agent framework achieved a better average score of 80.07% in IOU.

Assessment

Detection is a type of problem that straightforwardly can be formulated as the control or path-finding problem. Generally speaking, the states are defined as the pixel values that the agents observe at the currentstep, and the actions are defined as movements along the different axis of the environment plus some scalingfactors. That is why agent-based detection has the most considerable number of papers among all RL-related image detection tasks. Though related work in this field is still growing, some challenges exist.

Firstly, the generalizability and reproducibility of the agent-based methods still need to be further investigated. In practical application, the quality and local features of the image may vary by the noise anddistortion introduced in the imaging process. The trained agent may not always be capable of finding the target in clinical settings. Furthermore, the trigger of the termination state in the inference stage needs to be improved. The most commonly used criteria adopted now is the happening of oscillation. However, this may lead to a very ineffective convergence, and the agent might even be trapped at some local optimal point and never reach the actual destination. Real-time detection is another direction that has caused moreinterest in recent years. RL has proved its fast detection capability due to the non-exhaustive searching strategy. However, in some high dimensional data, 4D images (3D plus temporal), for example, the real- time detection and tracking still need more investigation. The last point is that the training process of the RL system, especially the multi-agent system, is very time-consuming, which may take days to weeks to train on even the best hardware platforms, let along the hyper parameter-tunning is also highly relied on the designer's experience. A summary of the works we reviewed in this section is given in Table 1.

Medical Image Segmentation

The key idea is to formulate this segmentation task as a control task by a simple Q-learning agent that decides the optimal local thresholds and the post-processing parameters. The quality of the segmentation is considered when designing the state. The segmentation threshold and size of the structuring elements are changed by taking a series of actions. Though simple as this initial research, the segmentation quality was acceptable while significantly reducing the required human interaction compared to the mainstream methods like active contour at that time.

Pre-locate the Segmentation Region

Most supervised-learning-based catheter segmentation methods require a large amount of well- annotated data. (Yang et al., 2020) proposed a semi-supervised pipeline shown in Figure 11 that first uses a DQN agent to allocate the coarse location of the catheter and then conducts patch-based segmentation byDual-UNet. The RL agent reduced the need for voxel-level annotation in the pre-allocation stage. The semi- supervised Dual-UNet exploited plenty of unlabeled images according to prediction hybrid constraints, thus improving the segmentation performance. The states are defined as the 3D observation patches, and the agent can update the states by moving the patch center point (x, y, z) along the x, y, and z-axis of the observation space. Like the landmark detection problems, the agent would give a negative reward if the patch moves away from the target; otherwise, a positive reward for moving toward and no reward if standing still. Compared to the state-of-the-art methods, this proposed pipeline requires much less computation time and achieves a minimum of 4% segmentation performance improvement measured by Dice Score.

Hyper parameters optimization

Instead of directly involved in the segmentation process, RL agents can also be applied to optimize the existing medical image segmentation pipelines (Bae et al., 2019; Qin et al., 2020; Yang et al., 2019). (Bae et al., 2019) used RL as the controller to automate the searching process of optimal neural architecture. The required search time and the computation power are significantly reduced by sharing the parameters while adopting a macro search strategy. Tested on the medical segmentation decathlon challenge, the authors assert that this optimized architecture outperformed the most advanced manually searched architectures.

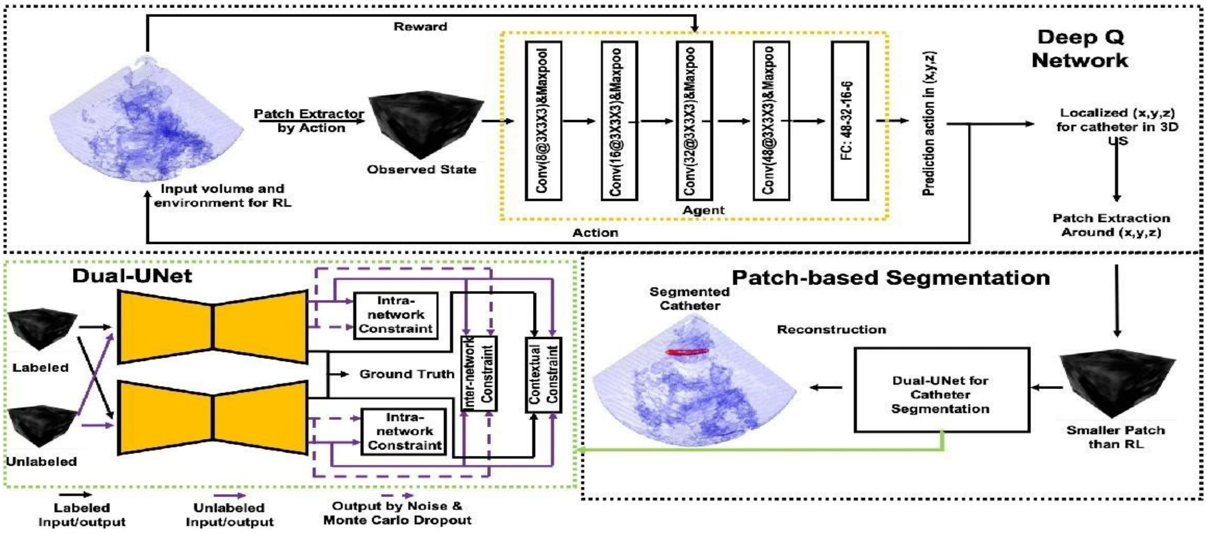

Figure 11: The semi-supervised DQN-driven catheter segmentation framework. Courtesy of (Yang et al., 2020).

Realizing the problem that some randomly augmented images might sometimes even harm the final segmentation performance, (Qin et al., 2020) implemented an automated end-to-end augmentation pipelineusing Dual DQN (DDQN) agent. By making the trails and saving the experiences, the agent would learn to determine the augmentation operations beneficial to the segmentation performance according to the fed- back Dice ratio. Twelve different basic actions would change the state to achieve augmentation. The state is defined as the extracted feature from U-Net. It is interesting to observe that horizontal flipping and cropping are two of the most useful operation.

(Yang et al., 2019) from NVIDIA integrated the highlights of the previous two reviewed papers. With an RNN-based controller, this research automates the design process of hyper- parameters and image augmentation to explore the maximum potential of the state-of-the- art models. The optimal policy is learned using the proximal policy optimization to decide the training parameters. Tested on the medical decathlon challenge tasks, the RL searched model and augmentation parameters have shown remarkable effectiveness and efficiency.

Segmentation as a Dynamic Process

Observing that many existing automated segmentation pipelines may often fail in real clinical applications, (Liao et al., 2020) implemented multi-agent reinforcement learning to interact with the users that can achieve an iteratively refined segmentation performance. This multi- agent strategy captures the dependence of the refinement steps and emphasizes the uncertainty of binary segmentation results in states$t$ indicates the current step. The actions will change the segmentation probability by an amount $a \in A$, where A is the action set. Furthermore, the voxel-wise reward is defined as $r^{(t)} = \chi^{(t-1)} - \chi^{(t)}$, where χ is the cross entropy between the label y_i probability p_i, to refine the segmentation more efficiently. The refined final segmentation result outperformed Min- Cut (Boykov & Kolmogorov, 2004), DeepGeoS(R- Net) (Wang et al., 2018), and Inter CNN (Bredell et al., 2018) on all the BRATS20015, MM- WHS, NCI- ICBI2013 datasets. Though published earlier than the (Liao et al., 2020) and adopted the older RL method to learn the policy, (Wang et al., 2013) incorporated not only the user's background knowledge but also their intentions. The proposed framework follows a "Show-Learn- Act" workflow, which reduces the required interactions while achieving context-specific and user-specific segmentation.

Assessment

Tackling the image segmentation problems using RL agents provides us with an effective way to furtheroptimize existing pipelines, overcome a limited number of training data, and interact with users to incorporate prior knowledge. Despite the novices of these methods, limitations still exist. The various definitions of states and actions may significantly influence the precision of the segmentation. In most works, the states are updated by a series of limited-number discrete actions to determine the final segmentation contours. Another problem is that the state design makes the agent only observe local or global information at a step. It would be interesting to see some methods in the future that can enable the agent to make these two pieces of information observable to the agent at the same time. A summary of the works we reviewed in this section is given in Table 2.

II. CONCLUSIONS

In this work, we have witnessed the success of some researchers' work that ingeniously turn the traditional image analysis tasks into RL-style behavioral or control problems. The basic concepts of reinforcement learning are first recapped, and a comprehensive analysis of applications of RL agents for different medical image analysis tasks was conducted in different sections. Under each section, the formulations of RL problems are discussed in detail from different angles. As the essential elements of theRL systems, the choice of algorithms, state, actions, and reward are highlighted in the table in Appendix

A. These RL-based methods provide us a way to think of the problems and create new paradigms for solving current obstacles. We hope that readers can find commonalities from these works, further understand the principles of reinforcement learning, and try to apply reinforcement learning in their futureresearch.

Disclosure

The authors are not aware of any affiliations, memberships, funding, or financial holds that might be perceived as affecting the objectivity of this review.

III. REFERENCES

[1]. Akrout, M., Farahmand, A.-m., Jarmain, T., & Abid, L. (2019). Improving skin condition classification with a visual symptom checker trained using reinforcement learning. International Conference on Medical Image Computing and Computer-Assisted Intervention,

[2]. Alansary, A., Oktay, O., Li, Y., Folgoc, L. L., Hou, B., Vaillant, G., Kamnitsas, K., Vlontzos, A., Glocker, B., Kainz, B., & Rueckert, D. (2019). Evaluating reinforcement learning agents for anatomical landmark detection [https://doi.org/10.1016/j.media.2019.02.007]. Medical image analysis, 53, 156-164.

[3]. Bae, W., Lee, S., Lee, Y., Park, B., Chung, M., & Jung, K.-H. (2019). Resource optimized neural architecture search for 3D medical image segmentation. International Conference on Medical

Image Computing and Computer-Assisted Intervention,

[4]. Boykov, Y., & Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence,26(9), 1124–1137-1124–1137.

[5]. Bredell, G., Tanner, C., & Konukoglu, E. (2018). Iterative Interaction Training for Segmentation Editing Networks. In Y. Shi, H.-I. Suk, & M. Liu, Machine Learning in Medical Imaging Cham.

[6]. Browning, J., Kornreich, M., Chow, A., Pawar, J., Zhang, L., Herzog, R., & Odry, B. L. (2021). Uncertainty Aware Deep Reinforcement Learning for Anatomical Landmark Detection in Medical Images. In M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, & C. Essert, Medical Image Computing and Computer Assisted Intervention – MICCAI 2021 Cham.

[7]. Ebrahimi, S., & Lim, G. J. (2021). A reinforcement learning approach for finding optimal policy of adaptive radiation therapy considering uncertain tumor biological response. Artificial Intelligence in Medicine, 121, 102193.

[8]. Eslami, M., Neuschaefer-Rube, C., & Serrurier, A. (2020). Automatic vocal tract landmark localization from midsagittal MRI data. Scientific reports, 10(1), 1468. https://doi.org/10.1038/s41598-020-58103-6

[9]. Fujimoto, S., Hoof, H., & Meger, D. (2018). Addressing function approximation error in actor-critic methods. International conference on machine learning,

[10]. Ghesu, F.-C., Georgescu, B., Zheng, Y., Grbic, S., Maier, A., Hornegger, J., & Comaniciu, D. (2019). Multi- Scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans [https://doi.org/10.1109/TPAMI.2017.2782687]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(1), 176-189.

[11]. Ghesu, F. C., Georgescu, B., Mansi, T., Neumann, D., Hornegger, J., & Comaniciu, D. (2016, 2016//). An Artificial Agent for Anatomical Landmark Detection in Medical Images. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016, Cham.

[12]. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio,

[13]. Y. (2014). Generative Adversarial Networks. In: arXiv.

[14]. Guan, Q., Chen, Y., Wei, Z., Heidari, A. A., Hu, H., Yang, X.-H., Zheng, J., Zhou, Q., Chen, H., & Chen, F. (2022). Medical image augmentation for lesion detection using a texture-constrained multichannel progressive GAN