Paper Reading No.14

Deep Reinforcement Learning for Active Breast Lesion Detection from DCE-MRI

Sheng Lian

October 2019

1 Brief Paper Intro

- ***Paper ref:*** MICCAI 2017, https://link.springer.com/chapter/10.1007/978-3-319-66179-7_76
- ***Authors:*** See Fig 1.

Gabriel Maicas[1(✉)], Gustavo Carneiro[1], Andrew P. Bradley[2], Jacinto C. Nascimento[3], and Ian Reid[1]

[1] School of Computer Science, ACVT, The University of Adelaide, Adelaide, Australia
gabriel.maicas@adelaide.edu.au
[2] School of ITEE, The University of Queensland, Brisbane, Australia
[3] Institute for Systems and Robotics, Instituto Superior Tecnico, Lisbon, Portugal

Figure 1: authors' brief intro.

- ***Paper summary:*** This paper proposed a deep reinforcement learning based method for automated detection of breast lesions from DCE-MRI. Such method can reduce the inference time for lesion detection and meanwhile retaining state-of-the-art accuracy.

- ***Reading motivation:*** Reinforcement learning, as a different task than supervised learning and unsupervised learning in the field of machine learning, is a field that I don't know much. In the all remaining 7 reading notes, I will read reinforcement learning-related papers, following Dr.Li's guide. So, I will read some RL-related application in medical image analysis (MIA), and some foundation blogs and papers on RL.

2 Methods

The detection process of breast lesions from DCE-MRI is shown in Fig 2. The proposed model comprises an artificial agent that automatically learns a policy that move and scale the bounding box towards the target (lesion).

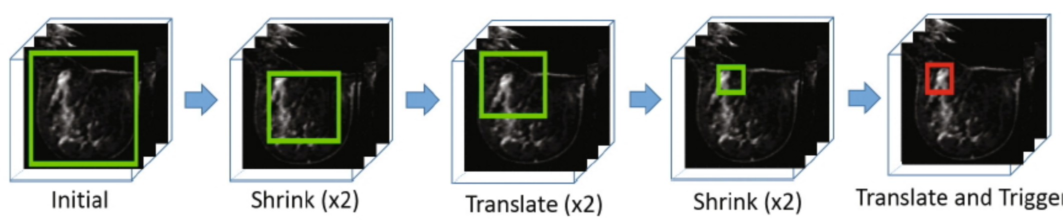


Figure 2: Example of the detection process of breast lesions from DCE-MRI with DQN. Depth transformation are not shown for simplicity..

In one step, the learning procedure is shown in Fig **??**. This figure is kind of brief. Given a observation (which contains the information of current bounding box and data), a ResNet is used to extract deep feature representation of the current bounding box. Then the DQN is followed to decide the next action, i.e., either to translate or scale the current bounding box or to trigger the end of the search process.

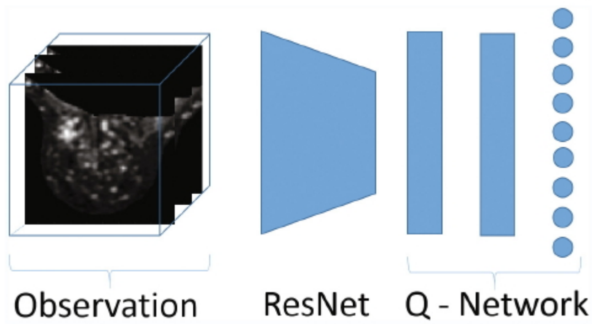


Figure 3: Block diagram of the proposed detection system.

2.1 Key issues in the proposed method

Here we will introduce some key issues in the proposed RL-based method.

2.1.1 Observation

Here, $\mathbf{o} = f(\mathbf{x}(\mathbf{b}))$ represent the observation, where $\mathbf{b} = [b_x, b_y, b_z, b_w, b_h, b_d] \in \mathbb{R}^6$ is the coordinates of top-left-front corner and lower-right-back corner of the current bounding box. f(.) denotes the ResNet.

2.1.2 Action

The action is denoted by $a \in \mathcal{A} = \{l_x^+, l_x^-, l_y^+, l_y^-, l_z^+, l_z^-, s^+, s^-, w\}$, where l,s,w represent translation, scale and trigger actions.

2.1.3 Reward

The reward when the agent chooses the action a = w to move from o to o' is defined by:

$$r\left(\mathbf{o}, a, \mathbf{o}'\right) := \begin{cases} +\eta, & \text{if } d\left(\mathbf{o}', \mathbf{s}\right) \geq \tau_w \\ -\eta, & \text{otherwise} \end{cases}$$

The threshold of Dice coefficient is been empirically set to be 0.2.

For the remaining actions in $\mathcal{A} \backslash \{w\}$, the rewards are defined by

$$r\left(\mathbf{o}, a, \mathbf{o}'\right) := \operatorname{sign}\left(d\left(\mathbf{o}', \mathbf{s}\right) - d(\mathbf{o}, \mathbf{s})\right)$$

2.1.4 Learning DQN

The training target here is to model a DQN to maximize cumulative future rewards, with the approximation of the following action-value function: $Q^*(\mathbf{o}, a) = \max_\pi \mathbb{E}\left[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots | \mathbf{o}_t = \mathbf{o}, a_t = a, \pi\right]$, And the loss function for modelling Q(o, a, θ) minimises the mean-squared error of the Bellman equation:

$$L_i\left(\theta_i\right) = \mathbb{E}_{\left(\mathbf{o}, a, r, \mathbf{o}'\right) \sim U(\mathcal{E})}\left[\left(r + \gamma \max_{a'} Q\left(\mathbf{o}', a'; \theta_i^-\right) - Q\left(\mathbf{o}, a; \theta_i\right)\right)^2\right]$$

The training of DQN is based on experience replay, which stores the data obtained by the exploration environment, and then randomly samples the parameters to update the parameters of the DQN. In the training process, the ϵ-greedy strategy is used to balance exploration and exploitation. This step add randomness to training process, which can improve generalisation. Finally, the ResNet, which produces the observation o = x(b), is trained to decide whether a random bounding box b contains a lesion.

The action to follow from the current observation is defined by:

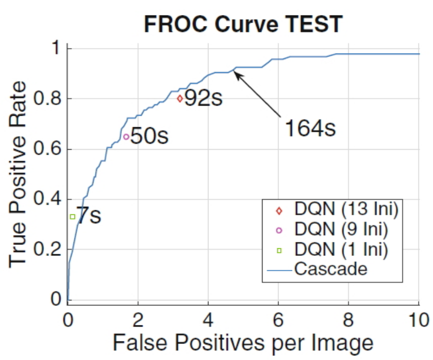
$$a^* = \arg\max_a Q\left(\mathbf{o}, a, \theta^*\right) \tag{1}$$

This inference is initialised with different bounding boxes at several locations, and it runs until it either finds the lesion (with the selection of the trigger action), or runs for a maximum number of 20 steps.

3 Experiments

The dataset used here is DCE-MRI data and T1-weighted data from 117 patients, 58for training, 59 for testing. The T1-weighted anatomical is used only to extract the breast region from the initial volume as a pre-processing stage.

The TPR, FPR, runtime are shown in Fig 4. I think here is a **mistake**. In the right side, it should be 'False positive per image', not 'FPR'.



	Test		
	TPR	**FPR**	**Time**
DQN (Ours)	**0.80**	**3.2**	**92 $\pm$ 21s**
Ms-C [6]	0.80	2.8	164$\pm$ 137s
SL [5]	1.00	4.50	$\mathcal{O}(60)$ m

Figure 4: FROC curve showing TPR vs FPI and run times for DQN and the multi-scale cascade (left) TPR, FPR and mean inference time per case (i.e. per patient) for each method (right). Note run time for Ms-C is constant over the FPI range.

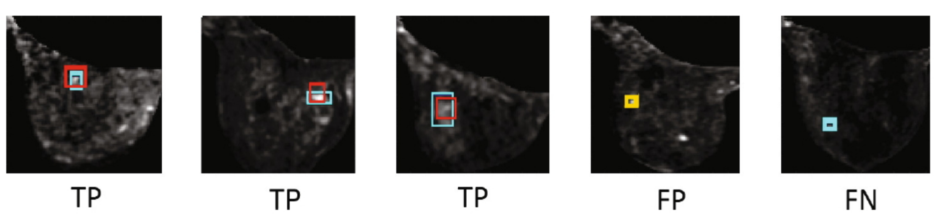


Figure 5: Examples of detected breast lesions. Cyan boxes indicate the ground truth, red boxes detections produced by our proposed method and yellow false positive detections..

4 Discussion

A DQN-based method is proposed for lesion detection. It shows comparable accuracy while with significantly reduced detection times. Using RL in such way is best used for initial positioning of candidate target or attention rectangle. But it is not suitable for accurate positioning, according to the example results.