Automatic Detection of the Retina in Optical Coherence Tomography using Deep Q Learning

Alex Cazañas-Gordón, Luís A. da Silva Cruz
Instituto de Telecomunicações
Department of Electrical and Computer Engineering
University of Coimbra, Coimbra, Portugal
acazanas@deec.uc.pt, lcruz@deec.uc.pt

***Abstract*—This study presents a novel approach to detecting the retina in optical coherence tomography (OCT) images using Deep Q learning. The proposed method uses an agent to extract contextual information from the input OCT to produce a tight-bounding box around the retina in a step-wise fashion. The detection task implements a decision process governed by a reinforcement learning strategy, where the agent takes actions and receives rewards according to their outcome. During the localization process, the agent learns the optimal set of actions to complete the detection task using a Q-network that estimates the value of the expected return of an action at any given step. Experiments on a test OCT dataset of 100 images showed that the proposed method accurately located the retina with a mean recall of 0.988 and a mean F1 score of 0.94.**

***Index Terms*—reinforcement learning, object detection, deep Q learning, optical coherence tomography**

I. Introduction

The retinal thickness is a widely used diagnostic marker of several eye-threatening conditions, including age-related macular degeneration (AMD), retinal vein occlusion (RVO), or diabetic macular edema (DME). Presently, the diagnosis of retinal pathology relies on several imaging modalities from which optic coherence tomography (OCT) stands out because of its highly detailed images of the retina and underlying structures. This imaging modality allows examining deep structures in the retina by acquiring cross-sectional images (B-scans) of the back of the eye. The segmentation of OCT B-scans is a fundamental step in measuring retinal thickness. Although modern OCT scanners include image processing tools that provide reasonably accurate segmentation of the retinal extent in healthy retinas [1], their performance degrades in the presence of degenerative diseases such as macular edema and retinal detachment [2], [3]. Retinal-pathology diagnosis relies heavily on the accurate localization of the retinal boundaries. However, OCT segmentation is labor-intensive, time-consuming and prone to inter-observer variability. Therefore, there is an unmet demand for automatic methods for localizing the retina in OCT B-scans.

Due to the wide variability of the appearance of retinal structures, several automatic methods approach the segmentation of OCT B-scans in a multi-stage fashion where the first step is identifying a region of interest (ROI) around the outermost layers of the retina [4]–[7]. This work focuses on developing an automatic algorithm that accurately locates the region that encloses the retinal boundaries. Specifically, the algorithm aims to demark the space between the inner-limiting membrane (ILM) and the retinal-pigmented epithelium (RPE). To this end, we propose a deep learning approach that implements a contextual-interactive decision process controlled by a deep reinforcement learning (DRL) strategy. The proposed method builds on the Deep-Q Network (DQN) algorithm [8] to learn a localization policy that determines the optimal sequence of actions to place a bounding box on top of the retina. The localization policy is optimized through trial and error using a localization agent, which acts upon the environment (a given input OCT B-scan) and obtains rewards according to the outcome of its actions. The agent follows a top-down search strategy in which a bounding box is successively reshaped until it closely surrounds the retina. The proposed method differs significantly from state-of-the-art localization approaches in that it does not enforce rigid search strategies such as sliding windows or fixed search paths. Instead, it localizes the object using a dynamic process, which progressively refines the focus of the search to find the target ROI. The main contribution of this work is the development of a novel localization method, which is able to produce precise localization results based on contextual information extracted dynamically from the environment.

II. Related work

To date, region-based convolutional neural networks (R-CNN) are the preferred approach to object detection and localization. RCNN-based detection has made considerable progress with the introduction of region proposal networks (RPN) and the development of Fast R-CNN [9], and Faster-R-CNN [10]. In the context of OCT image processing, R-CNNs have been used for the detection of the choroid [11], subretinal hemorrhage [12], vascular plaque [13], and DME [14]. Although R-CNN-based methods have shown promising results, their performance typically relies on a large number of object region proposals, which render them computationally expensive [15]. Moreover, like other supervised learning methods, R-CNNs are prone to overfitting. In particular, when complex

This work was funded by the Secretariat of Higher Education, Science, Technology and Innovation of the Republic of Ecuador and in part by the Fundação para a Ciência e a Tecnologia (FCT/MCTES), Portugal, through national funds and when applicable co-funded by EU funds under Projects UIDB/EEA/50008/2020, UIDP/50008/2020 and LA/P/0109/2020.

models are trained with little labeled data –a commonplace in the medical domain [16], [17]. DRL offers an alternative approach to learning where agents can learn with little or any labeled data. In related work, a number of studies have proposed DRL methods for object detection with promising results [18]–[20]. In addition, DQNs have been applied to detect anatomical landmarks [21], breast lesions [22], and pancreas [16]. However, to the best of our knowledge, no previous study has explored the application of DQN to the detection of anatomical structures on OCT B-scans.

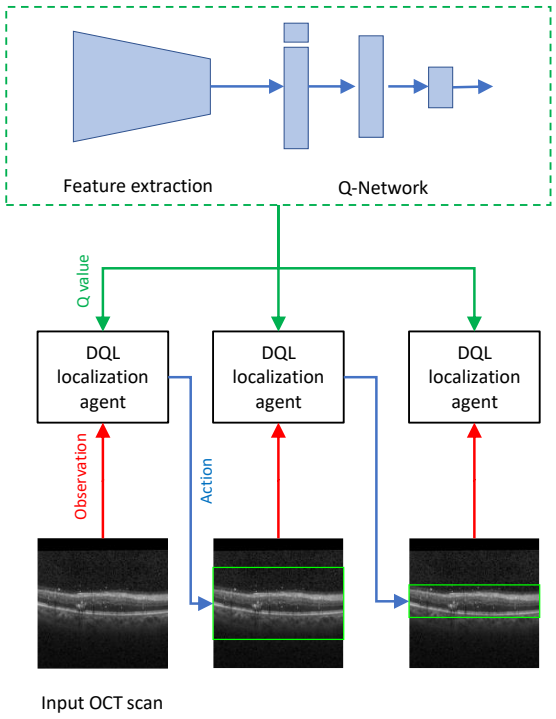

Fig. 1. Overview of the proposed localization approach. The localization agent takes contextual observations from the input OCT scan and transforms a bounding box seeking to maximizes the localization results.

III. Methods

The goal of the proposed algorithm is to produce a bounding box for enclosing the neurosensory retina in the input OCT B-scan. Considering the high variability of the location and shape of the retina, we adopted a localization approach based on the DQN algorithm.

A. Deep-Q Network

DQN is an extension of the classical reinforcement learning algorithm Q-learning [23] that uses deep neural networks. Q-learning sets an agent to interact with a dynamic environment and obtain rewards according to the outcome of its actions. At any given step, the agent performs an action a which is determined by its policy $\pi(a|s)$, where s is the current state of the environment. The outcome of the action a is a new state $s^{'}$ and a reward $r^{'}$. The goal of the agent is to learn an optimal policy so as to maximize the cumulative reward over the course of interactions with the environment. Central to the policy-optimization goal is the notion of a state-action value function, termed the Q-value function. The Q-value function of a policy π, $Q^{\pi}(s,a)$, measures the expected return obtained from taking action a at state s and then following the optimal policy. In the Q-learning algorithm the optimal long-term reward is estimated with the Bellman equation:

$$Q^{*}(s,a) = r + \gamma max_{a'} Q(s^{'}, a^{'}) \quad (1)$$

where r is the immediate reward, $max_{a'} Q(s^{'}, a^{'})$ represents the long-term reward and γ is the discount factor. In the DQN algorithm the agent uses a deep neural network (Q-Network) as a function approximator with a set of parameters θ to estimate the $Q(s,a|\theta)$ value. During training the agent tune the parameters of the Q-Network to maximize the cumulative reward.

B. DQN-based Localization

In the proposed method, a DQN agent is tasked to produce a tight-bounding box enclosing the retina. To that ned the agent takes a series of actions that transform an initial bounding box until it closely overlaps the ROI. At each step of the localization process, the DQN agent obtains contextual information from the environment and receives rewards according to how close the outcome of the bounding-box transformation matches the ROI. During the series of interactions with the environment, the DQN agent train a Q-network to learn the optimal localization policy. The detail of the parametrization of the state s, the action set A, and the reward r is presented below. Fig. 1 shows the overview of the proposed localization approach. An example of a typical action sequence is shown in Fig. 2.

1) Actions: As the retina occupies the entire width of any given B-scan, it is unnecessary to transform the box width but only its height. Accordingly, we restricted the box transformations to the vertical direction and defined two types of actions: 1) translation and 2) scaling, with two actions each (up and down). The action set A included five actions, four box transformations, and one trigger action to terminate the localization process. Any transformation action make a discrete change to the box representation, which is determined by its vertical coordinates: $b = [y_1, y_2]$. The box transformations are performed by adding or subtracting a fixed amount α to the box coordinates according to the desired effect. The trigger action terminates the action sequence on the current search and indicates that a suitable bounding box has been placed. Selecting the trigger action restarts the box representation and initiates a new search. Besides the trigger action, the localization process will also terminate upon reaching a predefined step limit. This step limit is an hyperparameter of the training process.

2) State: The state of the environment is fed to the Q-network, which predicts the Q value of four transformation actions with the current state. The action with the highest Q value is selected to generate the next bounding box. The

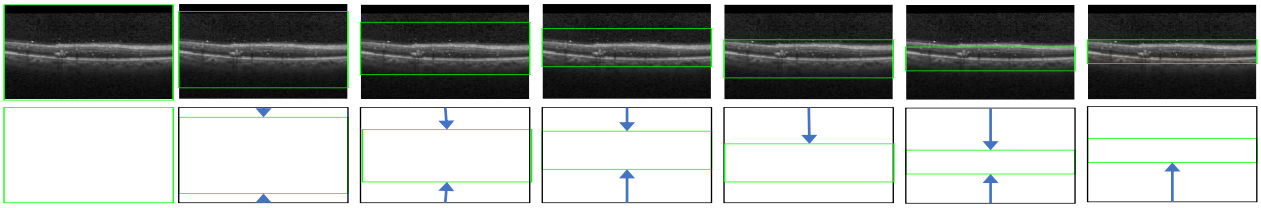

Fig. 2. Example of a typical localization sequence showing a series of transforming actions the DQN agent takes upon the bounding box. Green boxes illustrate the successive transformations of the bounding box. Blue arrows indicate the box transformation at each step.

state representation consists of two pieces of information, namely, the environment observation o and the action history h. The environment observation is a feature vector of the region visited by the agent. This feature vector is obtained by cropping the visited area from the OCT B-scan and feeding it to a convolutional neural network (CNN). At any given step, we stored the agent's experience e, which is a tuple containing the current state s, the action taken a, the reward r obtained as a result of the pair (s, a) and the next state $s^{'}$. The information stored in the agent experience allows the agent to replay past actions. Using a replay memory has been demonstrated to stabilize the search trajectory by making the agent escape repetitive cycles [24].

3) Reward: The reward r is determined by the outcome of the pair (s, a). The reward is granted according to the overlap between the bounding box B and the ground-truth G. To measure the overlap, we adopted the intersection-over-union (IoU) ratio, which is expressed as:

$$IoU(B, G) = \frac{B \cap G}{B \cup G} \tag{2}$$

The reward scheme is binary and it is determined according to the improvement of the IoU from one state to the next as follows:

$$r(s^{'}, s) = sign(IoU(B^{'}, G), IoU(B, G)) \tag{3}$$

The trigger action does not grant a reward, and it is selected whenever the IoU surpasses a predefined threshold τ.

IV. Experiments

To evaluate the proposed method, we used OCT B-scan images sourced from a publicly available dataset. We evaluated the F1-score and the detection precision using k-fold cross-validation. Details of the Q-network architecture and the training process are presented below.

A. Dataset

The OCT B-scan images used in this study were sourced from a publicly available dataset consisting of 10 OCT volumes from 10 healthy patients [25]. The scans were acquired with an SD-OCT Spectralis device (Heidelberg Engineering, Heidelberg, Germany). All volumes contain 10 B-scans, 496 pixels in height, and variable width (543 pixels to 644 pixels). The scans have lateral resolution 10-12 µm and axial resolution 3.87 µm. As a preprocessing step, we center-cropped the images along the longest axis to make them square with a side of 496 pixels. We applied a median filter with kernel size 3x3 pixels followed by a mean filter with kernel size 7x3 pixels to remove speckle noise. To improve contrast, we applied a power-law transformation [26] to the normalized pixel intensity. The size of the kernels and the value of the power-law transformation exponent were estimated empirically in prior work [4] to preserve the continuity of the retinal boundaries. Ground truths for training and testing were produced with the annotations of the ILM and the RPE layers. Lastly, we split the dataset into four disjointed partitions to perform cross-validation with k = 4 folds. In each fold, we evaluated the F1 score, and the precision, and reported the mean and standard deviation across four folds.

B. Q-network

The deep neural network used to estimate the Q value consisted of two blocks consisting of one fully connected layer and ReLu activation, followed by another fully connected layer. The network input is the current environment state, which comprises the feature vector of the region enclosed by the current-bounding box and the action history. Fig. 3 shows a block diagram of the Q-network architecture. To obtain the feature vector, we used an AlexNet [27] model pre-trained on the ImageNet dataset. Using the pre-trained model helped speeding up the learning process, as we only updated the parameters of the Q-network. Alternatively, to evaluate the impact of the transfer learning strategy, we added the feature extractor block of the AlexNet model to the Q-network and updated their weights from scratch. After initializing the Q-network parameters at random, the agent searches the environment for a suitable region to place the bounding box. The localization policy followed a ϵ-greedy strategy [28]. Accordingly, we set the agent to gradually drift from exploration to exploitation with ϵ-greedy exploration decay of $5\text{x}10^{-3}$. The discount factor γ was set to 0.9. The replay memory size was $5\text{x}10^{4}$ and the minibatch size 64. The agent was trained for 500 steps per episode, so we allowed 20 actions per image on average. For the box transformation parameter α, we chose a value of 1 since larger values made it harder to place the box in preliminary trials. The threshold τ of the trigger action was 0.8 in order to compel the agent to find a tight-bounding box for the ROI.

The networks were trained on a Windows 10 PC (CPU: Intel i7 8700K CPU @ 3.7 GHz - 6 cores, RAM: 32 GB)

equipped with a GPU NVIDIA GeForce GTX 1080 Ti with 11 GB RAM.

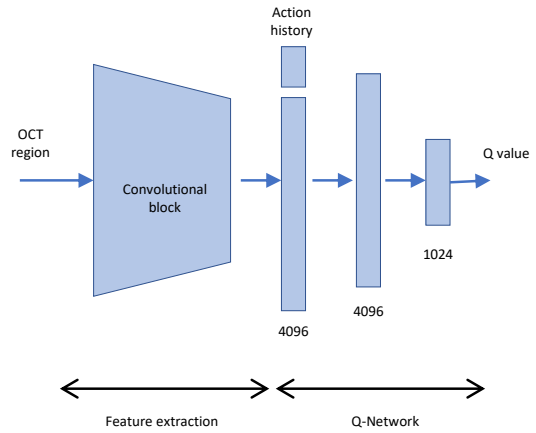

Fig. 3. Architecture of the Q-network

V. Results and Analysis

As shown in Table I there was no significant difference in performance between training the Q-network with and without the feature extractor block. However, we observed a faster convergence using the pre-trained AlexNet model. This result coincides with prior work using pre-trained models to obtain feature vectors for DQN-based object localization [16], [29]. The ϵ-greedy and replay memory strategies adopted during training were instrumental in stabilizing the learning process as well as lowering the oscillation between consecutive experience cycles. The training curve in Fig.4 was obtained with replay memory. By contrast, the curve in Fig.5 shows the training progress without replay memory. Even though the formulation of the reinforcement learning introduces randomness via the ϵ-greedy exploration strategy. We observed that the trained models showed consistent performance with a narrow standard deviation in 4-fold cross-validation results. Localization errors were produced in OCT B-scans showing a pronounced tilting (Fig. 6). We suspect that these errors occurred due to the relative under-representation of this type of image in the dataset, and should be addressed by increasing the diversity of the dataset. The proposed method matches the results of a Fast-RCNN model built with the same pre-trained AlexNet model used as a feature extractor for the DQN. The R-CNN was fine-tuned with the OCT B-scans for 30 epochs with the stochastic gradient descent using a learning rate of 10^{-4} and momentum = 0.9. As shown in Table I the performance of the proposed approach is comparable to that of a state-of-the-art deep learning localization method. Thus, it can be reliably used in OCT segmentation tasks to locate the region occupied by the retina.

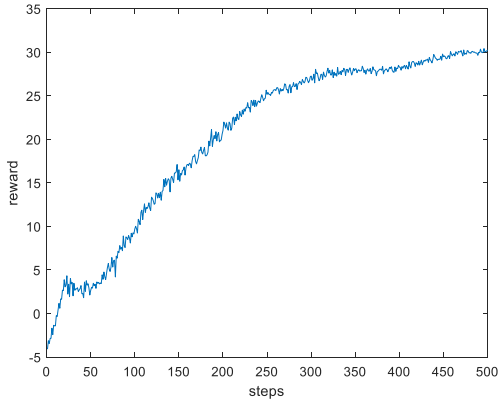

Fig. 4. Convergence curve of the training with replay memory

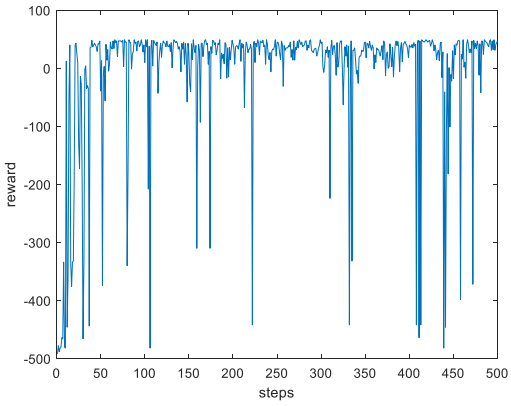

Fig. 5. Convergence curve of the training without replay memory

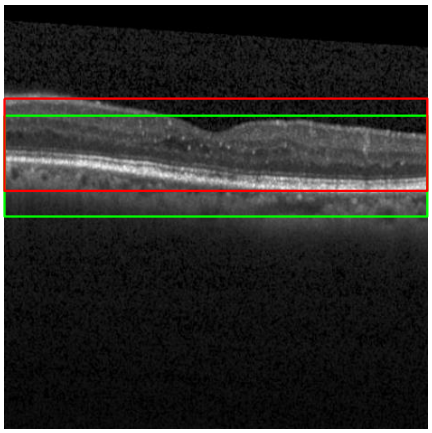

Fig. 6. Example of a localization error. Bounding box in red: ground truth, bounding box in green: DQN output

TABLE I
RESULTS OF THE PROPOSED METHOD AND COMPARATIVE OBJECT DETECTION METHOD. MEAN(STD) VALUES COMPUTED ACROSS 4 FOLDS.

Model	Recall	Precision	F1 score
Proposed DQN†	0.988(0.03)	0.911(0.12)	0.943(0.08)
Proposed DQN	**0.995(0.01)**	0.904(0.12)	0.942(0.08)
Fast-RCNN	0.980(0.02)	**0.989(0.01)**	**0.985(0.02)**

† pretrained AlexNet model used as feature extractor.

VI. CONCLUSIONS

This study proposes a novel method for locating the neurosensory retina in OCT B-scans using deep Q learning. The proposed approach addresses the need for automatic methods to annotate OCT B-scans, which ordinarily is a labor-intensive task. Results of the evaluation on a publicly available OCT dataset showed that the proposed method reached competitive performance on par with an RCNN model. The proposed method achieved a mean F1 score of 0.988 ± 0.03 and a mean recall of 0.943 ± 0.08 on the test OCT dataset. The proposed method introduces an effective alternative to state-of-the-art algorithms for the localization of diagnostic biomarkers in retinal OCT B-scans. In future work, we hope to extend the proposed method to other retinal pathology diagnosed with OCT, such as epiretinal membrane, macular fluid, macular holes, and drusen.

REFERENCES

[1] A. Aojula, S. P. Mollan, J. Horsburgh, A. Yiangou, K. A. Markey, J. L. Mitchell, W. J. Scotton, P. A. Keane, and A. J. Sinclair, "Segmentation error in spectral domain optical coherence tomography measures of the retinal nerve fibre layer thickness in idiopathic intracranial hypertension," *BMC ophthalmology*, vol. 17, no. 1, 2017. [Online]. Available: https://doi.org/10.1186/s12886-017-0652-7

[2] R. A. Alshareef, A. Goud, M. Mikhail, H. Saheb, H. K. Peguda, S. Dumpala, S. Rapole, and J. Chhablani, "Segmentation errors in macular ganglion cell analysis as determined by optical coherence tomography in eyes with macular pathology," *International journal of retina and vitreous*, vol. 3, no. 1, 2017. [Online]. Available: https://doi.org/10.1186/s40942-017-0078-7

[3] S. M. Waldstein, B. S. Gerendas, A. Montuoro, C. Simader, and U. Schmidt-Erfurth, "Quantitative comparison of macular segmentation performance using identical retinal regions across multiple spectral-domain optical coherence tomography instruments," *British Journal of Ophthalmology*, vol. 99, no. 6, pp. 794–800, 2015.

[4] A. Cazañas-Gordón, E. Parra-Mora, and L. A. D. S. Cruz, "Ensemble learning approach to retinal thickness assessment in optical coherence tomography," *IEEE Access*, vol. 9, pp. 67 349–67 363, 2021.

[5] J. Kugelman, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Automatic segmentation of OCT retinal boundaries using recurrent neural networks and graph search," *Biomedical optics express*, vol. 9, no. 11, pp. 5759–5777, 2018.

[6] M. Gende, J. De Moura, J. Novo, P. Charlón, and M. Ortega, "Automatic segmentation and intuitive visualisation of the epiretinal membrane in 3D OCT images using deep convolutional approaches," *IEEE Access*, vol. 9, pp. 75 993–76 004, 2021.

[7] E. Parra-Mora, A. Cazañas-Gordon, R. Proença, and L. A. da Silva Cruz, "Epiretinal membrane detection in optical coherence tomography retinal images using deep learning," *IEEE Access*, vol. 9, pp. 99 201–99 219, 2021.

[8] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[9] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[11] W. P. Hsia, S. L. Tse, C. J. Chang, and Y. L. Huang, "Automatic segmentation of choroid layer using deep learning on spectral domain optical coherence tomography," *Applied Sciences*, vol. 11, no. 12, p. 5488, 2021.

[12] M. Suchetha, N. S. Ganesh, R. Raman, and D. E. Dhas, "Region of interest-based predictive algorithm for subretinal hemorrhage detection using faster R-CNN," *Soft Computing*, vol. 25, no. 24, pp. 15 255–15 268, 2021.

[13] C.-Y. Sun, X.-J. Hong, S. Shi, Z.-Y. Shen, H.-D. Zhang, and L.-X. Zhou, "Cascade faster R-CNN detection for vulnerable plaques in OCT images," *IEEE Access*, vol. 9, pp. 24 697–24 704, 2021.

[14] J. Wu, Q. Zhang, M. Liu, Z. Xiao, F. Zhang, L. Geng, Y. Liu, and W. Wang, "Diabetic macular edema grading based on improved faster R-CNN and MD-ResNet," *Signal, Image and Video Processing*, vol. 15, no. 4, pp. 743–751, 2021.

[15] F. Navarro, A. Sekuboyina, D. Waldmannstetter, J. C. Peeken, S. E. Combs, and B. H. Menze, "Deep reinforcement learning for organ localization in CT," in *Proceedings of the Third Conference on Medical Imaging with Deep Learning*, ser. Proceedings of Machine Learning Research, T. Arbel, I. Ben Ayed, M. de Bruijne, M. Descoteaux, H. Lombaert, and C. Pal, Eds., vol. 121. PMLR, 06–08 Jul 2020, pp. 544–554.

[16] Y. Man, Y. Huang, J. Feng, X. Li, and F. Wu, "Deep Q learning driven CT pancreas segmentation with geometry-aware U-Net," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1971–1980, 2019.

[17] A. Cazañas-Gordón, E. Parra-Mora, and L. A. da Silva Cruz, "Evaluating transfer learning for macular fluid detection with limited data," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1348–1352.

[18] M. B. Bueno, X. G.-i. Nieto, F. Marqués, and J. Torres, "Hierarchical object detection with deep reinforcement learning," *Deep Learning for Image Processing Applications*, vol. 31, no. 164, p. 3, 2017.

[19] A. Pirinen and C. Sminchisescu, "Deep reinforcement learning of region proposal networks for object detection," in *proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6945–6954.

[20] B. Uzkent, C. Yeh, and S. Ermon, "Efficient object detection in large images using deep reinforcement learning," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2020, pp. 1824–1833.

[21] F.-C. Ghesu, B. Georgescu, Y. Zheng, S. Grbic, A. Maier, J. Hornegger, and D. Comaniciu, "Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 1, pp. 176–189, 2017.

[22] G. Maicas, G. Carneiro, A. P. Bradley, J. C. Nascimento, and I. Reid, "Deep reinforcement learning for active breast lesion detection from DCE-MRI," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2017, pp. 665–673.

[23] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3, pp. 279–292, 1992.

[24] R. Liu and J. Zou, "The effects of memory replay in reinforcement learning," in *2018 56th annual allerton conference on communication, control, and computing (Allerton)*. IEEE, 2018, pp. 478–485.

[25] J. Tian, B. Varga, G. M. Somfai, W.-H. Lee, W. E. Smiddy, and D. C. DeBuc, "Real-time automatic segmentation of optical coherence tomography volume data of the macular region," *PloS one*, vol. 10, no. 8, 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0133908

[26] R. C. Gonzalez, R. E. Woods *et al.*, *Digital image processing*. New York, NY, USA: Pearson, 2018, pp. 125–129.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

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

[29] J. C. Caicedo and S. Lazebnik, "Active object localization with deep reinforcement learning," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2488–2496.