Deep Q-learning for Active Recognition of GERMS: Baseline performance on a standardized dataset for active learning

Mohsen Malmir[1]
http://mplab.ucsd.edu/~mmalmir/
Karan Sikka[1]
http://mplab.ucsd.edu/~ksikka/
Deborah Forster[1]
dforster@ucsd.edu
Javier Movellan[2]
http://www.emotient.com/
Garrison W. Cottrell[3]
http://cseweb.ucsd.edu/~gary/

[1] Machine Perception Lab.
University of California San Diego,
San Diego, CA, USA

[2] Emotient, Inc.
4435 Eastgate Mall, Suite 320,
San Diego, CA, USA

[3] Computer Science and Engineering Dept.
University of California San Diego,
San Diego, CA, USA

Active object recognition (AOR) refers to problems in which an agent interacts with the world and controls its sensor parameters to maximize the speed and accuracy with which it recognizes objects. A wide range of approaches have been developed to re-position sensors or change the environment so that the new inputs to the system become less ambiguous [1, 2] with respect to goals such as 3D reconstruction, localization or recognition of objects. Many of the active object recognition methods are built around a specific hardware system, which makes the replication of their results very difficult. Other systems use off-the-shelf computer vision datasets, which include several views of objects captured by systematically changing object's orientation in the image. However, these datasets do not offer any active object recognition benchmark *per se*.

In this paper, we present and make publicly available the GERMS dataset (see Figure 1), that was specifically developed for active object recognition. The data collection procedure was motivated by the needs of the RUBI project, whose goal is to develop robots that interact with toddlers in early childhood education environments [4]. To collect data, we asked a set of human subjects to hand the GERM objects to RUBI in poses they considered natural. RUBI then pretends to examine the object by bringing it to its center of view and rotating the object. The background of the GERMS dataset was provided by a large screen TV displaying video scenes from the classroom in which RUBI operates, including toddlers and adults moving around.

Figure 1: Object set used in GERMS dataset. The objects represent human cell types, microbes and disease-related organisms.

We also propose an architecture (DQL) for AOR based on deep Q-learning (see Figure 2). To our knowledge, this is the first work employing deep Q-learning for active object recognition. An image is first transformed into a set of features using a DCNN borrowed from [3] which was trained on ImageNet. We add a softmax layer on top of this model to recognize GERMS objects; the output of this softmax layer is the belief over different GERMS objects given an image. This belief is combined with the accumulated belief from the previous images using Naive Bayes. This accumulated belief represents the *state* of the AOR system in each time step.

The accumulated belief is then transformed by the policy learning network into action values. This network is composed of two Rectified-Linear-Unit (ReLU) layers followed by a Linear-Unit (LU) layer. Each unit in the LU represents the action value for a given accumulated belief and one of the possible actions. In order to train this module, we implement the Q-learning iterative update:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\{R(s,a) + \gamma \max_{a^*} Q(s^*,a^*) - Q(s,a)\} \qquad (1)$$

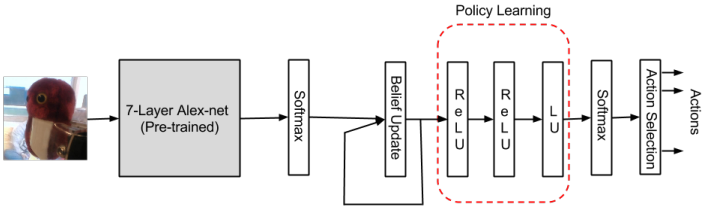

Figure 2: The proposed architecture for DQL.

into the following stochastic gradient descent weight update rule for the network:

$$W \leftarrow W - \lambda \left(R_t + \gamma \max_a Q(B_{t+1},a) - Q(B_t,a_t)\right) \frac{\partial}{\partial W} Q(B_t,a_t). \qquad (2)$$

Here, W is the weights of the policy learning network, $Q(s,a)$ is the action-value learned by the network for action a in state s, γ is the reward-discount factor and R_t is the reward at the tth time step.

The number of output units in the policy learning network is equal to the number of possible actions. Each output unit calculates the action value $Q(s,a)$ for one action a. We implemented a set of actions which rotate the robot's wrist from its *current position* by $\pm\pi/64$, $\pm\pi/32$, $\pm\pi/16$, $\pm\pi/8$, $\pm\pi/4$. The allowable range of rotation for both robot wrists is in $[0,\pi]$.

Table 1 shows the superior performance of the proposed DQL method compared to *random* and *sequential* action selection strategies. For detailed description of the dataset and the training algorithm, please refer to the paper.

Table 1: Number of steps required by the sequential, random and DQL policies to reach the same level of prediction accuracy on GERMS dataset.

Method \ Prediction Accuracy(%)	48	53	55	58	62	
Sequential	18	30	-	-	-	**Right Arm**
Random	2	4	6	10	-	
DQL	**1**	**2**	**2**	**3**	**10**	
Sequential	15	24	-	-	-	**Left Arm**
Random	3	10	18	-	-	
DQL	**1**	**3**	**3**	**7**	-	

[1] John Aloimonos, Isaac Weiss, and Amit Bandyopadhyay. Active vision. *International journal of computer vision*, 1(4):333–356, 1988.

[2] Ruzena Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8): 966–1005, 1988.

[3] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.

[4] Mohsen Malmir, Deborah Forster, Kendall Youngstrom, Lydia Morrison, and Javier R Movellan. Home alone: Social robots for digital ethnography of toddler behavior. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 762–768. IEEE, 2013.