An Artificial Agent for Anatomical Landmark Detection in Medical Images

Florin C. Ghesu[1,2(✉)], Bogdan Georgescu[1], Tommaso Mansi[1], Dominik Neumann[1], Joachim Hornegger[2], and Dorin Comaniciu[1]

[1] Medical Imaging Technologies, Siemens Healthineers, Princeton, NJ, USA
florin.c.ghesu@fau.de

[2] Pattern Recognition Lab, Friedrich-Alexander-Universität, Erlangen, Germany

Abstract. Fast and robust detection of anatomical structures or pathologies represents a fundamental task in medical image analysis. Most of the current solutions are however suboptimal and unconstrained by learning an appearance model and exhaustively scanning the space of parameters to detect a specific anatomical structure. In addition, typical feature computation or estimation of meta-parameters related to the appearance model or the search strategy, is based on local criteria or predefined approximation schemes. We propose a new learning method following a fundamentally different paradigm by simultaneously modeling both the object appearance and the parameter search strategy as a unified behavioral task for an artificial agent. The method combines the advantages of behavior learning achieved through reinforcement learning with effective hierarchical feature extraction achieved through deep learning. We show that given only a sequence of annotated images, the agent can automatically and strategically learn optimal paths that converge to the sought anatomical landmark location as opposed to exhaustively scanning the entire solution space. The method significantly outperforms state-of-the-art machine learning and deep learning approaches both in terms of accuracy and speed on 2D magnetic resonance images, 2D ultrasound and 3D CT images, achieving average detection errors of 1-2 pixels, while also recognizing the absence of an object from the image.

1 Introduction

At the core of artificial intelligence is the concept of knowledge-driven computational models which are able to emulate human intelligence. The textbook [8] defines intelligence as the ability of an individual or artificial entity to explore, learn and understand tasks, as opposed to following predefined solution steps.

Machine learning is a fundamental technique used in the context of medical image parsing. The robust detection, segmentation and tracking of the anatomy are essential in both the diagnostic and interventional suite, enabling real-time guidance, quantification and processing in the operating room. Typical

S. Ourselin et al. (Eds.): MICCAI 2016, Part III, LNCS 9902, pp. 229–237, 2016.
DOI: 10.1007/978-3-319-46726-9_27

machine learning models are learned from given data examples using suboptimal, handcrafted features and unconstrained optimization techniques. In addition, any method-related meta-parameters, e.g. ranges, scales, are hand-picked or tuned according to predefined criteria, also in state-of-the-art deep learning solutions [3,11]. As a result, such methods often suffer from computational limitations, sub-optimal parameter optimization or weak generalization due to overfitting, as a consequence of their inability to incorporate or discover intrinsic knowledge about the task at hand [1,5,6]. All aspects related to understanding the given problem and ensuring the generality of the algorithm are the responsibility of the engineer, while the machine, completely decoupled from this higher level of understanding, blindly executes the solution [8].

In this paper we make a step towards self-taught virtual agents for image understanding and demonstrate the new technique in the context of medical image parsing by formulating the landmark detection problem as a generic learning task for an artificial agent. Inspired by the work of Mnih *et al.* [7], we leverage state-of-the-art representation learning techniques through deep learning [1] and powerful solutions for generic behavior learning through reinforcement learning [10] to create a model encapsulating a cognitive-like learning process to discover strategies, i.e. optimal search paths for localizing arbitrary landmarks. In other words, we enable the machine to learn how to optimally search for a target as opposed to following time-consuming exhaustive search schemes. In parallel to our work, similar ideas have been exploited also in the context of 2D object detection [2].

2 Background

Building powerful artificial agents that can emulate or even surpass human performance at given tasks requires the use of an automatic, generic learning model inspired from human cognitive models [8]. The artificial agent needs to be equipped with at least two fundamental capabilities to achieve intelligence. At perceptual level is the automatic capturing and disentangling of high-dimensional signal data describing the environment, while on cognitive level is the ability to reach decisions and act upon the observed information [8]. Deep learning and reinforcement learning provide the tools to build such capabilities.

2.1 Deep Representation Learning

Inspired by the feed-forward type of information processing observable in the early visual cortex, the deep convolutional neural network (CNN) represents a powerful representation learning mechanism with an automated feature design, closely emulating the principles of the animal and human receptive fields [1]. The architecture is composed of hierarchical layers of translation-invariant convolutional filters based on local spatial correlations observable in images. Denoting the l-th convolutional filter kernel in the layer k by $\boldsymbol{w}^{(k,l)}$, we can write the representation map generated by this filter as: $o_{i,j} = \sigma((\boldsymbol{w}^{(k,l)} * \boldsymbol{x})_{i,j} + b^{(k,l)})$,

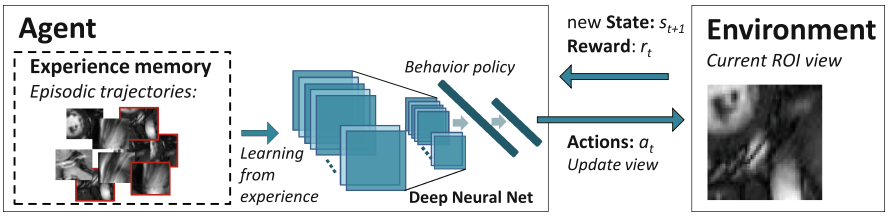


Fig. 1. System diagram showing the interaction of the artificial agent with the environment for landmark detection. The state s_t at time t is defined by the current view, given as an image window. The actions of the agent directly impact the environment, resulting in a new state and a quantitative feedback: (s_{t+1}, r_t). The experience memory stores the visited states, which are periodically sampled to learn the behavior policy.

where x denotes the representation map from the previous layer (used as input), (i, j) define the evaluation location of the filter and $b^{(k,l)}$ represents the neuron bias. The function σ represents the activation function used to synthesize the input information. In our experiments we use rectified linear unit activations (ReLU) given their excellent performance. In a supervised setup, i.e. given a set of independent observations as input patches $\boldsymbol{X}$ with corresponding value assignments $\boldsymbol{y}$, we can define the network response function as $\mathcal{R}(\,\cdot\,; \boldsymbol{w}, \boldsymbol{b})$ and use Maximum Likelihood Estimation to estimate the optimal network parameters: $\hat{\boldsymbol{w}}, \hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{w},\boldsymbol{b}} \|\mathcal{R}(\boldsymbol{X}; \boldsymbol{w}, \boldsymbol{b}) - \boldsymbol{y}\|_2^2$. We solve this optimization problem with a stochastic gradient descent (SGD) approach combined with the backpropagation algorithm to compute the network gradients.

2.2 Cognitive Modeling Using Reinforcement Learning

Reinforcement learning (RL) is a technique aimed at effectively describing learning as an end-to-end cognitive process [9]. A typical RL setting involves an artificial agent that can interact with an uncertain environment, thereby aiming to reach predefined goals. The agent can observe the state of the environment and choose to act on it, similar to a trial-and-error search [9], maximizing the future reward signal received as a supervised response from the environment (see Fig. 1). This reward-based decision process is modeled in RL theory as a *Markov Decision Process* (MDP) [9] $\mathcal{M} := (\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$, where: $\mathcal{S}$ represents a finite set of states over time, $\mathcal{A}$ represents a finite set of actions allowing the agent to interact with the environment, $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0; 1]$ is a stochastic transition function, where $\mathcal{T}_{s,a}^{s'}$ describes the probability of arriving in state s' after performing action a in state s, $\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ is a scalar reward function, where $\mathcal{R}_{s,a}^{s'}$ denotes the expected reward after a state transition, and γ is the discount factor controlling future versus immediate rewards.

Formally, the future discounted reward of an agent at time $\hat{t}$ can be written as $R_{\hat{t}} = \sum_{t=\hat{t}}^{T} \gamma^{t-\hat{t}} r_t$, with T marking the end of a learning episode and r_t defining the immediate reward the agent receives at time t. Especially in model-free reinforcement learning, the target is to find the optimal so called action-value

function, denoting the maximum expected future discounted reward when starting in state s and performing action a: $Q^*(s,a) = \max_\pi \mathbb{E}\left[R_t | s_t = s, a_t = a, \pi\right]$, where π is an action policy, in other words a probability distribution over actions in each given state. Once the optimal action-value function is estimated the optimal action policy, determining the behavior of the agent, can be directly computed in each state: $\forall s \in \mathcal{S} : \pi^*(s) = \arg\max_{a \in \mathcal{A}} Q^*(s,a)$. One important relation satisfied by the optimal action-value function Q^* is the Bellman optimality equation [9]. This is defined as:

$$Q^*(s,a) = \sum_{s'} \mathcal{T}_{s,a}^{s'} \left(\mathcal{R}_{s,a}^{s'} + \gamma \max_{a'} Q^*(s',a') \right) = \mathbb{E}_{s'} \left(r + \gamma \max_{a'} Q^*(s',a') \right), \quad (1)$$

where s' defines a possible state visited after s, a' the corresponding action and $r = R_{s,a}^{s'}$ represents a compact notation for the current, immediate reward. Viewed as an operator τ, the Bellman equation defines a contraction mapping. Strong theoretical results [9] show that by iteratively applying $Q_{i+1} = \tau(Q_i), \forall (s,a)$, the function Q_i converges to Q^* at infinity. This standard, model-based policy iteration approach is however not always feasible in practice. An alternative is the use of model-free temporal difference methods, typically Q-Learning [10], which exploit correlations of consecutive states. A step further towards a higher computational efficiency is the use of parametric functions to approximate the Q-function. Considering the expected non-linear structure of the Q-function [10], neural networks represent a potentially powerful solution for policy approximation [7]. In the following we leverage these techniques in an effort to make a step towards machine-driven intelligence for image parsing.

3 Proposed Method

We propose to formulate the image parsing problem as a deep-learning-driven behavior policy encoding automatic, intelligent paths in parametric space towards the correct solution. Let us consider the example of landmark detection. The optimal search policy in this case represents a trajectory in image space converging to the landmark location $p \in \mathbb{R}^d$ (d is the image dimensionality).

3.1 Agent Learning Model

As previously motivated, we model this new paradigm with an MDP $\mathcal{M}$. While the system dynamics $\mathcal{T}$ are implicitly modeled through our deep-learning-based policy approximation, the state space $\mathcal{S}$, the action space $\mathcal{A}$ and reward/feedback scheme $\mathcal{R}$ need to be explicitly designed:

– **States** describe the surrounding environment - in our context we model this as a focus of attention, a region of interest in the image with its center representing the current position of the agent.

– **Actions** denote the moves of the agent in the parametric space. We select a discrete action-scheme allowing the agent to move one pixel in all directions: *up, down, left, right* - corresponding to a shift of the image patch. This allows the agent to explore the entire image space.
– **Rewards** encode the supervised feedback received by the agent. Opposed to typical choices [7], we propose to follow more closely a standard human learning environment, where rewards are scaled according to the quality of a specific move. We select the reward to be δd, the supervised relative distance-change to the landmark location after executing a move.

3.2 Deep Reinforcement Learning for Image Parsing

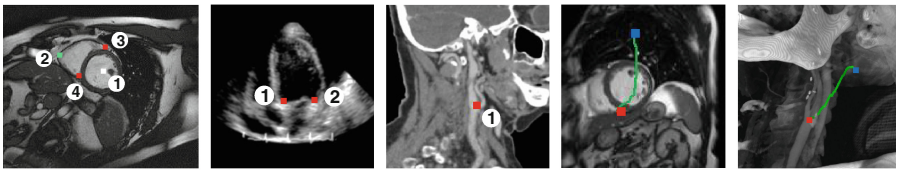
Given the model definition, the goal of the agent is to select actions by interacting with the environment in order to maximize cumulative future reward. The optimal behavior is defined by the optimal policy π^* and implicitly optimal action-value function Q^*. In this work we propose a model-free, temporal difference approach introduced in the context of game learning by Mnih *et al.* [7], using a deep CNN to approximate the optimal action-value function Q^*. Defining the parameters of a deep CNN as θ, we use this architecture as a generic, non-linear function approximator $Q(s, a; \theta) \approx Q^*(s, a)$ called deep Q network (DQN). A deep Q network can be trained in this context using an iterative approach to minimize the mean squared error based on the Bellman optimality criterion (see Eq. 1). At any learning iteration i, we can approximate the optimal expected target values using a set of reference parameters $\theta_i^{ref} := \theta_j$ from a previous iteration $j < i$: $y = r + \gamma \max_{a'} Q(s', a'; \theta_i^{ref})$. As such we obtain a sequence of well-defined optimization problems driving the evolution of the network parameters. The error function at each step i is defined as:

$$\hat{\theta}_i = \arg\min_{\theta_i} \mathbb{E}_{s,a,r,s'} \left[\left(y - Q(s, a; \theta_i) \right)^2 \right] + \mathbb{E}_{s,a,r} \left[\mathbb{V}_{s'}[y] \right]. \tag{2}$$

This is a standard, supervised setup for DL in both 2D and 3D (see Sect. 2).

Reference Update-Delay. Using a different network to compute the reference values for training brings robustness to the algorithm. In such a setup, changes to the current parameters θ_i and implicitly to the current approximator $Q(\cdot\,; \theta_i)$ cannot directly impact the reference output y, introducing an update-delay and thereby reducing the probability to diverge and oscillate in suboptimal regions of the optimization space [7].

Experience Replay. To ensure the robustness of the parameter updates and train more efficiently, we propose to use the concept of experience replay [4]. In experience replay, the agent stores a limited memory of previously visited states as a set of explored trajectories: $\mathcal{E} = [t_1, t_2, \cdots, t_P]$. This memory is constantly sampled randomly to generate mini-batches guiding the robust training of the CNN and implicitly of the agent behavior policy.



(a) 2D MRI (b) 2D US (c) 3D CT (d) 2D path (e) 3D path

Fig. 2. Figures depicting the landmarks considered in the experiments. Figure (a) shows the LV-center (1), RV-extreme (2) and the anterior / posterior RV-insertion points (3) / (4) in a short-axis cardiac MR image. Figure (b) highlights the mitral septal annulus (1) and the mitral lateral annulus points (2) in a cardiac ultrasound image and figure (c) the right carotid artery bifurcation (1) in a head-neck CT scan. Figures (d) and (e) depict trajectories/optimal paths followed by the agent for detection, blue denotes the random starting point, red the groundtruth and green the optimal path. (Color figure online)

4 Experiments

Accurate landmark detection is a fundamental prerequisite for medical image analysis. We developed a research prototype to demonstrate the performance of the proposed approach on this type of application for 2D magnetic resonance (MR), ultrasound (US) and 3D computed tomography (CT) images.

4.1 Datasets

We use three datasets containing 891 short-axis view MR images from 338 patients, 1186 cardiac ultrasound apical four-chamber view images from 361 patients and 455 head-neck CT scans from 455 patients. The landmarks selected for testing are presented in Fig. 2. The train/cross-validation/test dataset split is performed randomly at patient level, for the MR dataset 711/90/90 images, for the US dataset 991/99/96 images and for the CT dataset 341/56/58 images. The results on the MR dataset are compared to the state-of-the art results achieved in [5,6] with methods combining context modeling with machine-learning for robust landmark detection. Please note that we use the same dataset as [5,6], but a different train/test split. On the CT dataset we compare to [11], a state-of-the-art deep learning solution combined with exhaustive hypotheses scanning. Here we use the same dataset and data split. In terms of preprocessing we resample the images to isotropic resolution, 2 mm in 2D and 1 mm in 3D.

4.2 Learning How to Find Landmarks

The learning occurs in episodes in which the agent explores random paths in random training images, constantly updating the experience memory and implicitly the search policy modeled by the deep CNN. Based on the cross-validation set we systematically select the meta-parameters and number of training rounds

Table 1. Table showing the detection error on the test sets with superior results highlighted in bold. The error is quantified as the distance to the ground-truth, measured in *mm*. With * we signify that the results are reported on the same dataset, but on a different training/test data-split than ours.

	Detection error [mm]											
	2D-MRI								2D-US		3D-CT	
	LV-center		RV-ext		RV-post			RV-ant	M-sep	M-lat	Bifurc.	
	Our	[6]	Our	[5]	Our	[6]	[5]	Our	Our	Our	Our	[11]
Mean	**1.8**	6.2*	**4.9**	8.4*	**2.2**	7.9*	5.9*	**3.7**	**1.3**	**1.6**	**1.8**	2.6
Median	**1.7**	5.4*	**4.2**	5.9*	**1.8**	4.7*	3.9*	**3.0**	**1.2**	**1.3**	**0.8**	1.2
STD	**2.2**	4.0*	**3.6**	16.5*	**1.5**	11.5*	16.0*	**2.3**	**0.8**	**1.4**	**2.9**	5.0

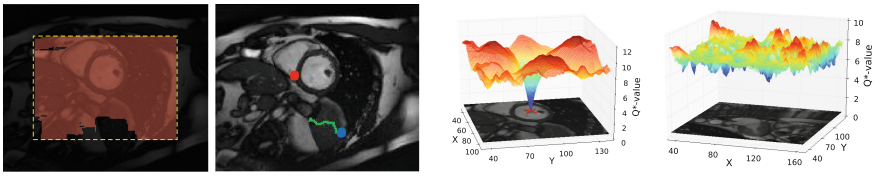
following a grid search: $\gamma = 0.9$, replay memory size $P = 100000$, learning rate $\eta = 0.00025$ and ROI 60^2 pixels, respectively 26^3 voxels. The network topology is composed of 3 convolution+pooling layers followed by 3 fully-connected layers with dropout. We emphasize that except for the adaptation of the CNN to use 3D kernels on 3D data, the meta-parameters are kept fixed for all experiments.

Policy Evaluation. During the evaluation the agent starts in a random state and follows the optimal policy with *no* knowledge about the groundtruth, navigating through the image space until an oscillation occurs - an infinite loop between two neighboring states, indicating the location of the sought landmark. The location is considered a high-confidence landmark detection if the expected reward from this location $\max_a Q^*(s_{target}, a) < 1$, i.e. the agent is closer than one pixel. This means the policy is consistent, rejecting the possibility of a local optimum and giving a powerful confidence measure about the detection. Table 1 shows the results on the test sets for all modalities and landmarks.

Object not in the Image? Using this property we not only detect diverging trajectories, but can also recognize if the landmark is not contained in the image. For example we evaluated trained agents on 100 long-axis cardiac MR images from different patients, observing that in such cases the oscillation occurs at points where $\max_a Q^*(s_{target}, a) > 4$. This suggests the ability of our algorithm to detect when the anatomical landmark is absent. (see Fig. 3(c–d)).

Convergence. We observed in random test images that typically more than 90 % of the possible start points converge to the solution (see Fig. 3(a–b)).

Speed Performance. While typical state-of-the-art methods [3,11] exhaustively scan solution hypotheses in large 2D or 3D spaces, the agent follows a simple path (see Fig. 2(d-e)). The average speed-up to scanning with a similar network (see for example [11]) is around **80× in 2D** and **3100× in 3D**. The very fast detection in 3D in less than 0.05 seconds highlights the potential of this technology for real-time applications, such as tracking of anatomical objects.



(a) Conv. map (b) Failed case (c) Object present (d) Object missing

Fig. 3. Figure (a) highlights in transparent red all the starting positions converging to the landmark location (the border is due to the window-based search). Figure (b) shows an example of a failed case. Figures (c) and (d) visualize the optimal action-value function Q^* for two images, the latter not containing the landmark. For this image there is no clear global minimum, indicating the absence of the landmark.

5 Conclusion

In conclusion, in this paper we presented a new learning paradigm in the context of medical image parsing, training intelligent agents that overcome the limitations of standard machine learning approaches. Based on a Q-Learning inspired framework, we used state-of-the-art deep learning techniques to directly approximate the optimal behavior of the agent in a trial-and-error environment. We evaluated our approach on various landmarks from different image modalities showing that the agent can automatically discover and efficiently evaluate strategies for landmark detection at high accuracy.

References

1. Bengio, Y., Courville, A.C., Vincent, P.: Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives. CoRR abs/1206.5538 (2012)
2. Caicedo, J.C., Lazebnik, S.: Active object localization with deep reinforcement learning. In: IEEE ICCV, pp. 2488–2496 (2015)
3. Ghesu, F.C., Krubasik, E., Georgescu, B., Singh, V., Zheng, Y., Hornegger, J., Comaniciu, D.: Marginal space deep learning: efficient architecture for volumetric image parsing. IEEE TMI **35**(5), 1217–1228 (2016)
4. Lin, L.J.: Reinforcement Learning for Robots Using Neural Networks. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, USA (1992)
5. Lu, X., Georgescu, B., Jolly, M.-P., Guehring, J., Young, A., Cowan, B., Littmann, A., Comaniciu, D.: Cardiac anchoring in MRI through context modeling. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 383–390. Springer, Heidelberg (2010)
6. Lu, X., Jolly, M.-P.: Discriminative context modeling using auxiliary markers for LV landmark detection from a single MR image. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 2012. LNCS, vol. 7746, pp. 105–114. Springer, Heidelberg (2013)
7. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D.: Human-level control through deep reinforcement learning. Nature **518**(7540), 529–533 (2015)

8. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach, 2nd edn. Pearson Education, Upper Saddle River (2003)
9. Sutton, R.S., Barto, A.G.: Introduction to Reinforcement Learning, 1st edn. MIT Press, Cambridge (1998)
10. Watkins, C.J.C.H., Dayan, P.: Q-learning. Mach. Learn. **8**(3), 279–292 (1992)
11. 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, Heidelberg (2015). doi:10.1007/978-3-319-24553-9_69