Program Digest, 2019 IEEE International Ultrasonics Symposium (IUS)
Glasgow, Scotland, October 6-9, 2019

Automatic Ultrasound Guidance Based on Deep Reinforcement Learning

Piotr Jarosik*, Marcin Lewandowski‡§
*Department of Information and Computational Science,
‡Laboratory of Professional Electronics,
Institute of Fundamental Technological Research PAS,
Warsaw, Poland;
§us4us Ltd., Warsaw, Poland;
email: pjarosik@ippt.pan.pl, mlew@ippt.pan.pl

***Abstract*—Ultrasound is becoming the modality of choice for everyday medical diagnosis, due to its mobility and decreasing price. As the availability of ultrasound diagnostic devices for untrained users grows, appropriate guidance becomes desirable. This kind of support could be provided by a *software agent*, who easily adapts to new conditions, and whose role is to instruct the user on how to obtain optimal settings of the imaging system during an examination. In this work, we verified the feasibility of implementing and training such an agent for ultrasound, taking the deep reinforcement learning approach. The tasks it was given were to find the optimal position of the transducer's focal point (FP task) and to find an appropriate scanning plane (PP task). The ultrasound environment consisted of a linear-array transducer acquiring information from a tissue phantom with cysts forming an object-of-interest (OOI). The environment was simulated in the Field-II software. The agent could perform the following actions: move the position of the probe to the left/right, move focal depth upwards/downwards, rotate the probe clockwise/counter-clockwise, or do not move. Additional noise was applied to the current probe setting. The only observations the agent received were B-mode frames. The agent acted according to stochastic policy modeled by a deep convolutional neural network, and was trained using the vanilla policy gradient update algorithm. After the training, the agent's ability to accurately locate the position of the focal depth and scanning plane improved. Our preliminary results confirmed that deep reinforcement learning can be applied to the ultrasound environment.**

***Index Terms*—ultrasound guidance, reinforcement learning, deep learning**

I. Introduction

The growing availability of ultrasound devices moves us towards the possibility to decrease waiting time, and thus increase the number of examinations crucial to the subject. However, in most cases tests should be performed by an experienced radiologist, who may not be available immediately. For instance, in cardio-oncology [1], [2], it may be desirable to perform accurate echocardiography for cancerous subjects before applying a treatment causing cardiotoxicity. In special cases a non-trained examiner can be guided by a computer system [3] or an experienced person available remotely [4]. In this work, we seek an automated way of supporting novice ultrasound users.

Automatic ultrasound guidance could be provided by a computer agent previously trained by interacting with the modality and the subject. Recent developments in the machine learning field suggests that this kind of agent could be trained using deep reinforcement learning algorithms. This approach was already successfully verified in the field of artificial intelligence and robotics. For example, Mnih et al. [5] showed that it is possible to train a computer program to play Atari 2600 computer games with a deep variant of Q-learning algorithm, by providing only raw pixels and current score to an agent. Lillicrap et al. [6] experimented with continuous action space environments, like car driving or dexterous manipulation. Zhang et al. [7] used Deep Q Network to train an agent to perform robotic manipulation tasks using solely camera frames.

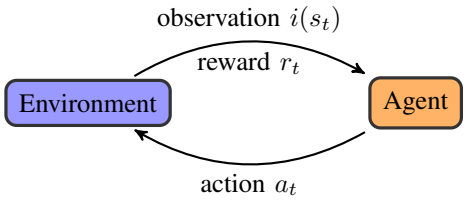


Fig. 1. The interaction between an agent and an environment.

In this work we investigate the feasibility of using the deep reinforcement learning approach in ultrasound guidance. Here, we train and evaluate an agent on a *toy environment* as presented in Figure 2. The agent observes B-mode frame only and manipulates the current transducer settings. The objective of the agent is to find and maintain a setting that provides the best quality of an ultrasound image sequence.

II. Method

In the reinforcement learning scenario an *agent* interacts with some *environment* in the following manner: the agent perceives some *observation* of the state s_t of the environment, *acts* (performs action a_t) according to its own inner policy and receives some numerical *reward* r_t (Figure 1). The agent's objective is to maximize its expected cumulative reward (a *return*) obtained over one run (an *episode*) [8], [9].

In this work, we consider an environment model comprised of a phantom with a simple object of interest inside. An

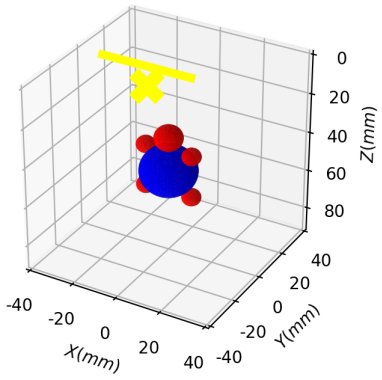
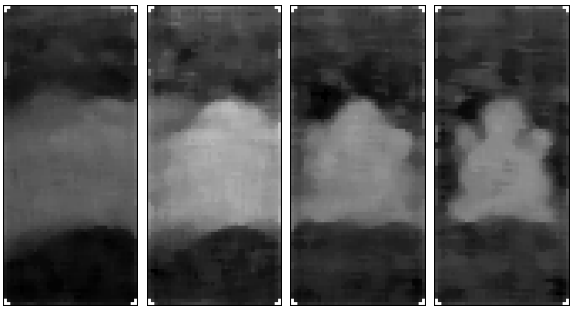


Fig. 2. A visualization of the environment. The OOI is composed of 6 balls: a *corpus* (the blue ball), a *head* (the larger red ball at the top of the corpus), and *legs* (the smaller 4 red balls with a 90 degree angle between them). An ultrasound transducer is applied to the top of the phantom (yellow line), and can change its lateral position, focus depth (yellow cross) and the scanning plane.



(a) (b) (c) (d)

Fig. 3. Example observations obtained by an agent during one episode of locating appropriate position of the focal point. At each time step the only provided observation is an ultrasound image $i(s_t)$.

ultrasound transducer is placed on the side of the phantom as presented in Figure 2. The only observation an agent receives is a B-mode image obtained at current position of the probe. The only action the agent can perform is to change the transducer settings: its lateral position, depth of the focal point and scanning plane rotation. In consequence, the environment may change its state, and the agent receives a new observation and reward. The *better the new image is* (in terms of the object's presentation), the greater the reward the agent obtains. The goal of the agent is to maximize the cumulative quality of a sequence of B-mode frames obtained over one examination. In the following subsections, we describe all these concepts in more detail.

A. Environment

The environment includes a cuboid phantom with an object inside, as presented in Figure 2. The object is composed of 6 balls, which are positioned and scaled so that they resemble a *corpus*, a *head* and 4 *legs*. The location of the object in the phantom is fixed and does not change between episodes and time steps. A linear transducer is located at the top of the phantom and can be: shifted by $\pm\Delta x$ millimeters along the line $M = \{(x, y, z) : y = 0, z = 0\}$, rotated $\pm\Delta\alpha$ degrees around the axis $Z = \{(x, y, z) : x = 0, y = 0\}$, and its focal point can be moved by $\pm\Delta z$. The transducer cannot be moved outside the phantom, nor can its focal point. Random displacements (a *noise*) of the probe were applied with the probability p_n. An occurrence of the noise in time step t means that after performing the selected action the transducer will be moved to left/right $\pm c_x\Delta x$ or rotated $\pm c_\alpha\Delta\alpha$ degrees additionally. For details, like the values of the initial state of the environment, please refer to the Table I.

TABLE I
ENVIRONMENT PARAMETERS

Parameter	**Value**
number of time steps per episode	16
phantom (width, height, depth) (mm)	$(80, 80, 90)$
transducer width (mm)	40
object location (x_o, y_o, z_o) (mm)	$(0, 0, 0)$
object angle α_o (mm)	0
transducer initial location (x, y, z) (mm)	$(x, 0, 0)$ FP: drawn from $[-20, 20]$ PP: drawn from $[-15, 15]$ divisible by Δx
transducer initial angle α (degrees)	FP : 0 PP: drawn from [45, 90] divisible by $\Delta\alpha$
transducer initial focal depth z_d (mm)	10
step size Δx (mm)	5
rotation angle $\Delta\alpha$ (degrees)	15
output image	40×90 pixels, grayscale
c_x	FP:drawn from$\{1, 2\}$,PP:1
c_α	PP: 1

The only observation the agent receives is a 2D B-mode image (Figure 3). The linear-array transducer and an ultrasound wave propagation (through the collection of point scatterers) were simulated in the Field II software [10], [11]. The resultant radio-frequency echo signal has undergone further processing comprised of: envelope detection (using Hilbert transform), log compression, dynamic range adjustment, interpolation to the output resolution, and a median filter.

In this work, we consider the following two separate *tasks*: given the initial (random) state of the environment, find (1) the appropriate position of the focal point or find (2) the appropriate scanning plane. For each of these tasks, we describe below an action space and reward function employed to train the agent.

1) Find the appropriate Focal Point (FP) Task: the environment starts with the transducer randomly located on the line M. The objective is to find the position of the transducer and the depth of the focal point, which gives the best *quality* image.

In FP task, the action space was discrete and consisted of

the following 5 operations:

- move the transducer to the left/right Δx millimeters,
- move the transducer's focal point depth up/down $\Delta z = \Delta x$ millimeters,
- do nothing (*NOP*).

The reward function was equal to the negative L_1 distance between the location of the transducer's focal point (x_t, z_t) and the center of the object $(x_{o,t}, z_{o,t})$ in step t:

$$R_t = -(|x_t - x_{o,t}| + |z_t - z_{o,t}|) \quad (1)$$

In other words, the L_1 distance between the position of the focal point and the position of the object's center was treated as a measure of image degradation (a *cost* $C(s_t)$): the further the focal point was from the object, the worse the image obtained. The objective could thus be formulated as minimizing the expected cumulative cost, or maximizing its negative version.

2) Find the appropriate Probe's Plane (PP) Task: the environment starts with the transducer, randomly rotated around the axis Z, randomly located on the line M and focused at the depth of the object's center. The objective is to find the position and the scanning plane of the transducer, which gives the best *quality* image.

In PP task, the action space was discrete and consisted of following 5 operations:

- move the transducer to the left/right Δx millimeters,
- rotate transducer $\Delta\alpha$ degrees counter-clockwise/clockwise,
- do nothing (*NOP*).

The reward in step t was equal to:

$$R_t = -(|x_t - x_{o,t}| + |sin(\alpha_t - \alpha_{o,t})|) \quad (2)$$

B. Agent

A simple model-free, on-policy gradient update algorithm [9] (a.k.a. *Vanilla Policy Gradient*) was used to train the agent to act in the environment and to perform tasks described in section II-A. The observation $i(s_t)$ is provided as an input to the policy π and value V functions, which were implemented as neural networks with 2D convolution-max pooling blocks, followed by a fully connected output layer; actions are sampled from the categorical distribution with parameters determined by the policy's neural network output. Batch normalization was applied after each convolutional layer [12]. Policy was trained by maximizing expected undiscounted return, and value function by minimizing mean-squared error with the discounted sum of rewards [9]. Selected hyperparameters of the training procedure can be found in Table II.

III. Results

The experiment results were averaged over 4 runs with different random number generator seeds. For both tasks, the return increased with the number of policy update iterations (Figure 4). The average distances of the transducer's focal point and scanning plane from object's setting for the PP task are presented in Figure 5. Both values decreased with time as expected. Increasing the likelihood of noise impacts the trend

TABLE II
Training hyperparameters

Parameter	Value
π learning rate	10^{-4}
V learning rate	10^{-3}
GAE λ (see [13])	FP: 0.95 PP: 0.97
discount factor γ	0.99
time steps per epoch	64

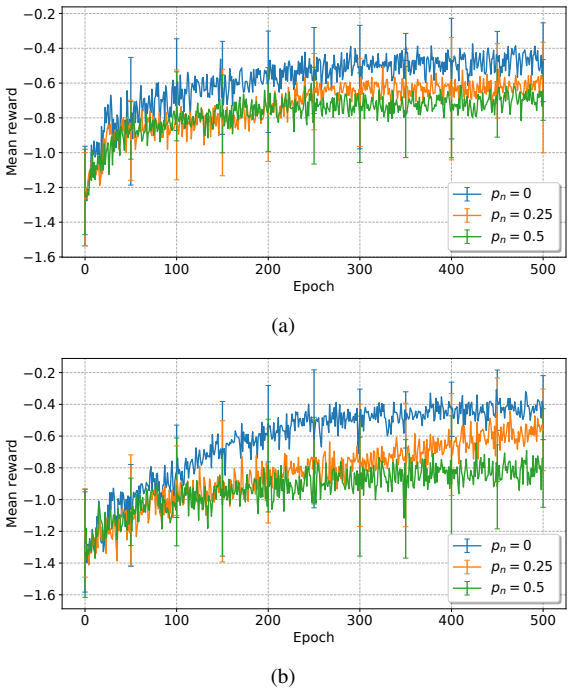
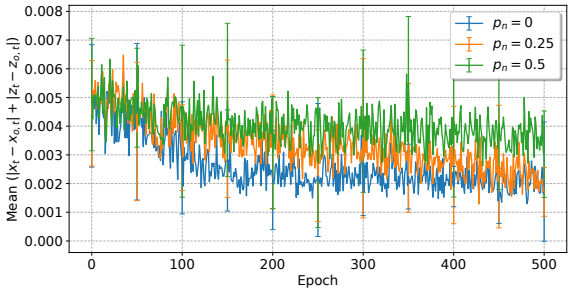


Fig. 4. Average reward after a given number of policy updates ($\pm$ std. dev. after each 50 epochs) for (a) FP task (also a *negative* distance from the the object) and (b) PP task, for different probabilities of applying the noise. Values were averaged across 4 runs with different random seeds.

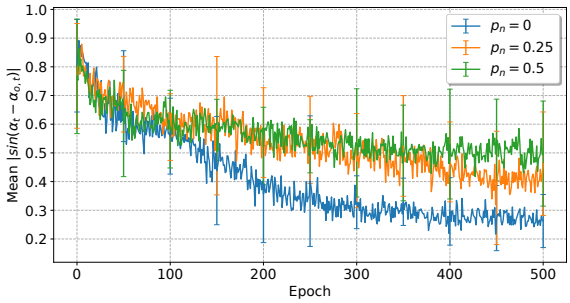
of a learning curve, especially for the PP task, but the agent's performance improves with time in all cases. While analyzing each episode individually we noticed that, in the FP task, the agent sometimes misses the target depth when moving the focal point from the top to the bottom (images obtained at and below the target depth have similar quality). In the PP task, the agent sometimes rotates the probe in a non-optimal direction (but still achieves the appropriate plane, due to the symmetry of the object).

IV. Conclusion

In this work, we proposed the deep reinforcement learning approach for the ultrasound guidance and verified its feasibility on a simple toy problem of improving the quality of an ultrasound image sequence. The results show that it is possible to train an agent to complete the task by changing the transducer setting, basing it only on the current B-mode frame. The implementation of the ultrasound environment was open-



(a)



(b)

Fig. 5. PP task results: (a) a distance from the object's center and (b) sinus of the angle between transducer's and object's planes. Values were averaged across 4 runs with different random seeds.

sourced and made available publicly for further experiments [14].

The approach described in this work can be extended in the future; here are some thoughts to consider. First, in a real-world case, the true location of the object of interest is usually unknown, thus a reward function should not be based on this information in general. A trained radiologist, an end user or some pre-trained model could score an image in non-simulated environments (*a quality of an image* can be a subjective term). The possibility to set other imaging settings/parameters by the agent, could also be investigated. Our experiments were also limited to a single, simple, on-policy, *vanilla* policy gradient algorithm – it would certainly be worth to evaluate the performance of other reinforcement learning methods in the ultrasound environment.

References

[1] L. Venneri, G. Zoppellaro, and R. S. Khattar, "Cardio-oncology: the role of advanced echocardiography in cancer patients," *Expert review of cardiovascular therapy*, vol. 16, no. 4, pp. 249–258, 2018.

[2] C. M. Larsen and S. L. Mulvagh, "Cardio-oncology: what you need to know now for clinical practice and echocardiography," *Echo research and practice*, vol. 4, no. 1, pp. R33–R41, 2017.

[3] A. Yen, A. K. Chaudhry, C. Wang, X. Tang, H. Hong, N. Poilvert, and D. Liang, "Innovative ultrasound technologies echogps and autoef help novices perform efficient and accurate echocardiographic monitoring in cancer patients," *Journal of the American College of Cardiology*, vol. 73, no. 9 Supplement 1, p. 1502, 2019.

[4] W. Conard, "Butterfly network announces the world's first augmented reality telemedicine technology," *Globenewswire*. [Online]. Available: https://www.globenewswire.com/news-release/2018/03/25/1452541/0/en/Butterfly-Network-Announces-the-World-s-First-Augmented-Reality-Telemedicine-Technology.html

[5] 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, p. 529, 2015.

[6] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[7] F. Zhang, J. Leitner, M. Milford, B. Upcroft, and P. Corke, "Towards vision-based deep reinforcement learning for robotic motion control," *arXiv preprint arXiv:1511.03791*, 2015.

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

[9] J. Achiam. (2019) Spinning up in deep reinforcement learning. [Online]. Available: https://spinningup.openai.com/

[10] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10th Nordic Baltic Conference on Biomedical Imaging, vol. 4, supplement 1, part 1: 351–353*. Citeseer, 1996.

[11] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 39, no. 2, pp. 262–267, 1992.

[12] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[13] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," *arXiv preprint arXiv:1506.02438*, 2015.

[14] P. Jarosik. (2019) Reinforcement learning for an ultrasound: source code repository. [Online]. Available: https://github.com/pjarosik/rlus