# Localization of Lumbar and Thoracic Vertebrae in 3D CT Datasets by Combining Deep Reinforcement Learning with Imitation Learning

Sankaran Iyer[1] Arcot Sowmya[1] Alan Blair[1] Christopher White[3]
Laughlin Dawes[2] Daniel Moses[2]

[1] School of Computer Science and Engineering,
University of New South Wales, Australia
[2] Department of Medical Imaging,
Prince of Wales Hospital, NSW, Australia
[3] Department of Endocrinology and Metabolism,
Prince of Wales Hospital, NSW, Australia

**Technical Report**
**UNSW-CSE-TR-201901**
**September 2019**

UNSW
SYDNEY

School of Computer Science and Engineering
The University of New South Wales
Sydney 2052, Australia

## Abstract

Landmark detection and 3D localization are often an important step in the analysis of medical images. This task, however, is challenging, due to the natural variability of human anatomical structures. We present a novel approach to lumbar and thoracic vertebrae localization by combining Deep Reinforcement Learning with Imitation Learning. The method involves navigating a 3D bounding box to the target landmark, followed by adjustment of the bounding box dimensions to enclose the region of interest (ROI). Two different 3D Convolutional Neural Networks (CNN) are used, one for learning the navigation in the coordinate directions, the other for predicting the bounding box dimensions. The algorithm is a modification of Deep Reinforcement Learning (Deep Q Networks), with the random search for navigation replaced by guiding the movement in an optimal coordinate direction using Imitation Learning. To improve the accuracy of detection, three different architectures for CNNs are used and the combined results provided to the next stage for analysis. Threefold cross validation is used to evaluate localization performance on two separate datasets, one each for the lumbar and thoracic spine. The method achieves mean 3D Jaccard Index of 76.96%(Dice Coefficient 85.92%) on the lumbar spine dataset after training on 115 Computed Tomography (CT) images and testing on 29. The corresponding figures for the thoracic spine are Jacquard index of 74.39% (Dice Coefficient 85%) after training on 105 and testing on 27. The results for this new approach are promising and the method is applicable for localization of any ROI in a 3D dataset.

**Keywords:**3D Localisation, Deep Reinforcement Learning, Imitation Learning, Convolution Neural Networks, Intersection Over Union

# 1 Introduction

Clinical examination of back pain and vertebral fractures requires analysis of the thoracic and lumbar spine regions. Computed Tomography (CT) datasets are more suited for this task as they provide better visualization of bone structures. Automated computer aided analysis of spine datasets requires localization of the Region of Interest (ROI) as a first step. Despite current approaches based on Geometric structures, Machine Learning and Deep Learning, processing of datasets in 3D continues to be a challenge. This paper presents a method based on Deep Reinforcement Learning and Imitation learning to address this problem.

# 2 Related Work

Traditional methods for vertebrae detection require prior knowledge of vertebrae locations, usually obtained from manual identification or statistical modelling, and detectors based on Geometric structures [1, 2] and the Generalized Hough Transform [3] have been used. Machine learning methods have also been employed along with feature de-scriptors: Support Vector Machines[4], Regression Forests [5], Adaboost [6] and Deformable Parts Model [7]. Many methods require a priori knowledge of vertebrae visibility and are therefore difficult to evaluate. The target ROI were also different, and the evaluation metrics were not consistent. Recent papers on vertebrae localization employ deep learning techniques using Deep Feed Forward neural networks [8], Multi-layered Perceptron (MLP) [9] and 3D CNN [10] . But these methods are focused on localization of vertebrae only. This paper is motivated by the idea of finding a general approach to 3D bounding box localization of any ROI, drawing inspiration from [11] for detecting 3D land-marks using Deep Reinforcement Learning. We propose a novel method, combining Deep Reinforcement learning and Imitation learning to localize lumbar and thoracic spine from CT datasets ,

## 2.1 Contributions:

The main contributions include a methodology to:

i navigate to the ROI by combining Deep Reinforcement Learning and Imitation Learning

ii predict the bounding box sizes upon reaching the ROI

iii finetune the bounding box sizes

# 3 Background and Proposed Work

Deep Reinforcement Learning has seen major successes in recent times [12] by combining the representation power of CNNs with Reinforcement learning. Using a Markov Decision Process (MDP), an artificial agent can be trained to achieve an intended goal. At any given time, an agent in a state $s_t$ selects an action $a_t$ from action space A based on policy $\pi(a_t|s_t)$ which represents the agents behaviour. The agent is taken to state $s_{t+1}$ and receives a reward $r_t$. In an episodic problem, this process continues till a terminal state is reached. The expected return at the end of the episode is the discounted accumulated reward with $\gamma$ being the discount factor:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \gamma \epsilon (0, 1] \tag{3.1}$$

The goal is to maximize this reward. The expected future discounted rewards for a given action a in a state s for a policy $\pi$ is known as Q value and is given by

$$Q^{\pi}(s, a) = E[R_(t)|s_t = s, a_t = a] \tag{3.2}$$

The optimal value function at any given state s for an action a is $Q^*$. Q learning involves updating the action value as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \big[r + \gamma max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\big] \tag{3.3}$$

where $\alpha$ is the learning rate. The agent has two choices in a state:

i explore by selecting a random action with probability $\epsilon$

ii exploit using already gained knowledge by choosing an action with the maximum Q value

After each episode, the state is reset to the initial value and the process repeated until the Q value converges.

Deep Reinforcement Learning has been used in bounding box object localization in 2D datasets [13]. However, bounding box localization in 3D has remained a challenge due to high computational resource requirements. Recently Deep Reinforcement Learning has been used for detection of anatomical landmarks in 3D CT datasets [11] by training an artificial agent to navigate from a random starting point towards the landmark and learning to move in the correct direction in the three coordinates. Learning is achieved by performing random searches, which is more appropriate for applications like gaming where there is a need to determine strategies for navigation. For landmark detection, it is more relevant and less complex for the agent to be trained in a guided manner. A navigation strategy to locate a landmark, as illustrated in Figure 3.1, to move in the coordinate direction at maximum distance from the current location to the center of the ground truth should suffice. We posit that it is appropriate to use a guided approach based on Imitation Learning.

Imitation learning is a paradigm for an agent to acquire skills by observing an expert [14]. Unlike Reinforcement learning, where the task of associating a state to actions is learned over several iterations, Imitation learning associate states with actions chosen by the expert. This converts the task to one of supervised learning of the mapping from states to expert actions.

The approach to localization in this work is to surround the ROI (lumbar/thoracic vertebrae) with 3D bounding boxes by combining the Deep Q learning algorithm [12] with Imitation learning when searching for an ROI from a predefined starting point in the image.

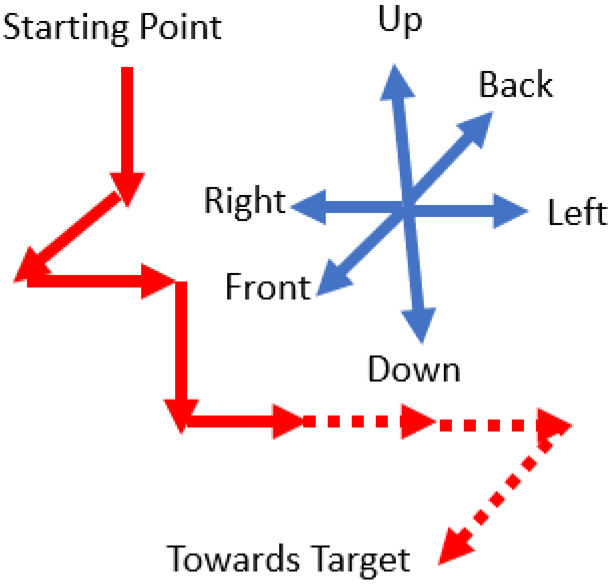


Figure 3.1: Red arrows show a navigation trajectory. Blue arrows show possible directions at each state

# 4 Method

## 4.1 Dataset and pre-processing

The dataset for vertebral analysis was provided by the Prince of Wales Hospital, Randwick, NSW, Australia in an anonymized form after ethics clearance. The CT datasets were acquired in a staged manner for both chest and abdominal regions. Abdominal datasets are required for lumbar spine analysis and chest datasets for thoracic spine analysis. The datasets were manually annotated and verified by the radiologist to identify the two diagonally opposite corner points of a 3D bounding box around the ROI using ITK-SNAP. The annotation process using ITK-SNAP in the three planes is illustrated in Figure 4.1.

## 4.2 Algorithm for Training

The algorithm involves training two networks:

i the first network to navigate a preselected bounding box to the centre of the ROI

ii the second network to predict the actual size of the bounding box surrounding the ROI

**Algorithm 1** Training by combining Deep Reinforcement Learning with Imitation Learning for ROI Detection

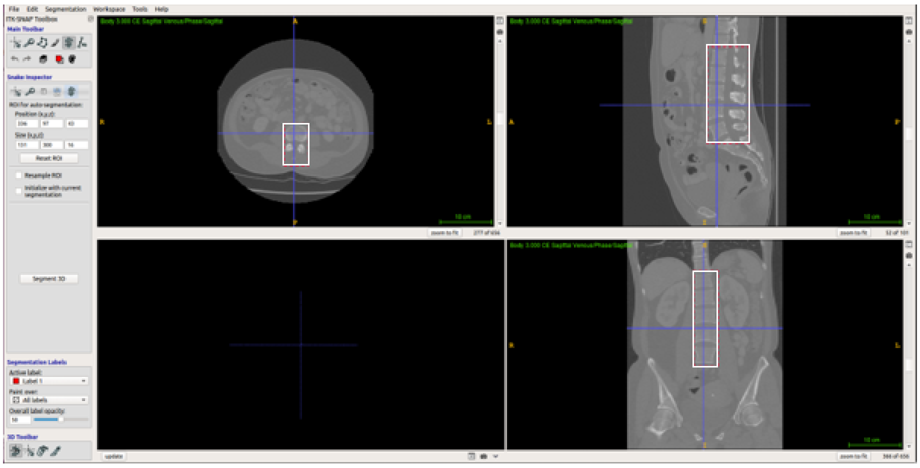


Figure 4.1: 3D Bounding Box annotation using ITK-SNAP shown by white boxes in 3 planes.

```
      Input: CT chest  abdominal 3D datasets
   Output: Policy function from which policy and action are selected for each
region within a bounding box, Bounding Box function that predicts the actual
bounding box coordinate sizes for each region within a bounding box
initialize Policy replay memory D
initialize Bounding Box replay memory B
initialize action-value function Q with random weights
for episodes from 1 to M
  for each a range of starting points
    for each dataset selected at random from the training set
      set a bounding box with mean coordinate dimension
      from the training set at a predefined starting point = $s_1$
      for steps from 1 to $N$
        following $\epsilon-$greedy policy select an action
          $$a_t = \left\{ \begin{array}{c} Imitation\ action\ with\ probability\ \epsilon \\ argmax_a\ Q(s_t, a)\ otherwise \\ Correction\ is\ applied\ by\ Imitation\ function \\ if\ predicted\ direction\ is\ away\ from\ Target \end{array} \right\}$$
        execute action a$_i$ $to\ shift\ image\ to\ s_{t+1}$
        $store\ transition\ s_t, a_t\ in\ D$
        $calculate\ the\ IOU\ of\ s_t\ with\ the\ ground\ truth$
        if $it\ exceeds\ a\ threshold\ level\ store\ s_t,$
          $ground\ truth\ bounding\ box\ coordinate\ sizes\ in\ B$
          $set\ s_t = s_{t+1}$
          if $bounding\ box\ centre\ has\ reached\ ground\ truth\ centre$
            $set\ a_t = Terminate$
            $store\ resulting\ transitions\ in\ D\ and\ B$
          break
        end for
```

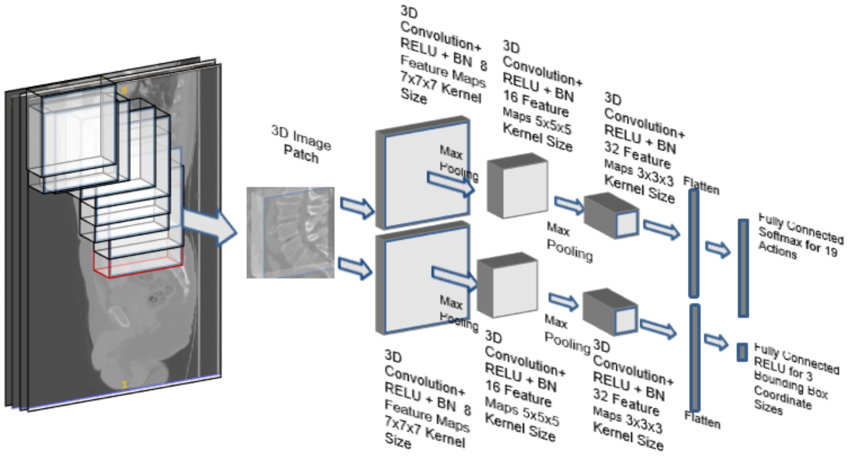

3D Image Patch
3D Convolution+ RELU + BN 8 Feature Maps 7x7x7 Kernel Size
Max Pooling
3D Convolution+ RELU + BN 16 Feature Maps 5x5x5 Kernel Size
Max Pooling
3D Convolution+ RELU + BN 32 Feature Maps 3x3x3 Kernel Size
Flatten
Fully Connected Softmax for 19 Actions
3D Convolution+ RELU + BN 8 Feature Maps 7x7x7 Kernel Size
Max Pooling
3D Convolution+ RELU + BN 16 Feature Maps 5x5x5 Kernel Size
Max Pooling
3D Convolution+ RELU + BN 32 Feature Maps 3x3x3 Kernel Size
Flatten
Fully Connected RELU for 3 Bounding Box Coordinate Sizes

Figure 4.2: Navigation of the bounding box to the Region of Interest (ROI). The red bounding box is the target ROI

```
            $select\ random\ samples\ from\ D\ and\ train\ Policy$
              $network\ with\ loss\ =\ mean\ square\ error\ between$
              $actual\ and\ predicted$
            $select\ random\ samples\ from\ B\ and\ train\ Bounding$
              $Box\ network\ with\ loss\ =\ mean\ square\ error$
              $between\ actual\ and\ predicted$
        end for
    end for
end for
```

The algorithm is illustrated in Figure 4.2 in and the pseudo code in Algorithm 1 The upper network in Figure 4.2 is the Policy network that is trained to predict the coordinate direction of shift (action) for an image region bounded by an initial preselected bounding box. In each coordinate direction, three levels of movement in the positive and negative directions are permitted.The three levels are coarse equalling a displacement by 25 voxels, fine by 10 voxels and very fine by 1 voxel respectively. In each coordinate direction, 3 levels of movement of the bounding box in both positive and negative directions requires 6 actions. In all 18 actions are possible for the 3 coordinates. The Imitation function in Algorithm 1 returns the action, which is the coordinate direction with maximum distance from the ground truth centre. It also corrects predictions deviating from the intended course. The appropriate level (i.e. coarse, fine or very fine) is selected based on the distance. The starting point for the first navigation trajectory is set at 20% of the coordinate sizes to eliminate margins and extract meaningful information from the datasets. Thereafter the network is trained by shifting the initial starting point by 25 voxels in the three coordinate directions till 80% of coordinate sizes is reached, to help the model recover from unfamiliar locations.

A final action called Terminate is used to indicate that the ground truth centre has been reached. Thus, the network should predict 19 possible actions in all.

The Policy network is made up of three 3D Convolution Layers together with Batch Normalization and RELU activation. The kernel size of first, second and third Convolution layers are 7x7x7, 5x5x5 and 3x3x3 respectively. The convolution layers are followed first by a fully connected layer and then by a softmax layer for 19 possible actions. The network takes as input the data within the bounding box shrunk by half. The convolution layers are followed first by a fully connected layer and then by a softmax layer for 19 possible actions.

To evaluate a localization, we use Intersection over Union (IOU) of the predict-ed bounding box with the ground truth. We use standard 50% threshold level for IOU for detection, as used in ImageNet and Regions with CNN for 2D bounding boxes [15, 16]. IOU is also known as Jaccard Index. We also report Dice Coefficient (DC) which is the ratio of twice the intersection over sum of the volumes ground truth and predicted bounding boxes.

The lower network in Figure 4.2 is the Bounding Box network, trained to predict the three coordinate sizes of the ROI. As the preselected bounding box is navigated, those regions whose IOU exceed a threshold level are stored along with the ground truth sizes for training the Bounding Box network. The latter is made up of three 3D Convolution Layers together with Batch Normalization and RELU activation. The kernel size of first, second and third Convolution layers are 7x7x7, 5x5x5 and 3x3x3 respectively. The network takes as input the data within the bounding box shrunk by half. The convolution layers are followed first by a fully connected layer and then by a RELU layer for 3 coordinate sizes.

In order to improve overall performance, it was decided to train two other architectures of CNNs besides the above and the predicted bounding boxes using all three stages are provided to the next stage for analysis. The architecture of the second model consists of 6 convolution layers. The first 2 layers have kernel size of 7x7x7, followed by 2 convolution layers with kernel size 5x5x5 and the final 2 convolution layers having kernel size 3x3x3. Each convolution layer is followed by Batch normalization. Max Pooling is added after the second and fourth layer. . The third model has a convolution layer with 9x9x9 kernel and a batch normalization preceding the architecture in the first model.

### 4.3 Testing Mode

In the testing mode there is no Imitation Learning involved during the navigation stage. Each test image was simply run for 25 steps which was found to be sufficient to reach the ROI. The search also terminates when a Terminate action is triggered or when a loop is detected between the states.

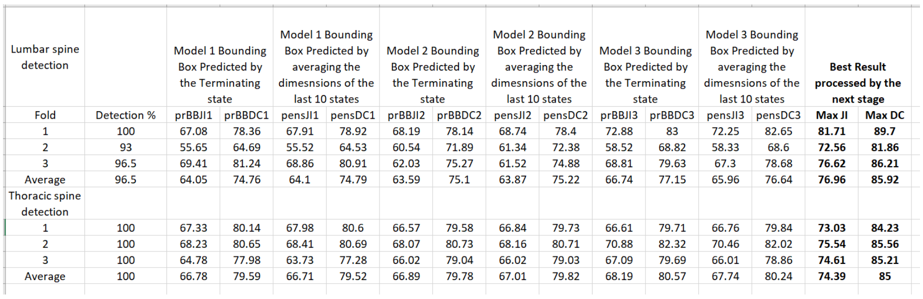
The bounding box prediction was run on all the steps and two different methods were used to predict the size:

i the predicted size of the Terminating state

ii the mean size of the predicted bounding boxes of the last 10 states

## 5 Experiments and Results

The training was run for 25 episodes on a Keras/Tensorflow platform. The learning rate was set to 0.00001. The starting point for navigation was set at



| Lumbar spine detection | | Model 1 Bounding Box Predicted by the Terminating state | | Model 1 Bounding Box Predicted by averaging the dimesnsions of the last 10 states | | Model 2 Bounding Box Predicted by the Terminating state | | Model 2 Bounding Box Predicted by averaging the dimesnsions of the last 10 states | | Model 3 Bounding Box Predicted by the Terminating state | | Model 3 Bounding Box Predicted by averaging the dimesnsions of the last 10 states | | **Best Result processed by the next stage** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fold | Detection % | prBBJI1 | prBBDC1 | pensJI1 | pensDC1 | prBBJI2 | prBBDC2 | pensJI2 | pensDC2 | prBBJI3 | prBBDC3 | pensJI3 | pensDC3 | **Max JI** | **Max DC** |
| 1 | 100 | 67.08 | 78.36 | 67.91 | 78.92 | 68.19 | 78.14 | 68.74 | 78.4 | 72.88 | 83 | 72.25 | 82.65 | **81.71** | **89.7** |
| 2 | 93 | 55.65 | 64.69 | 55.52 | 64.53 | 60.54 | 71.89 | 61.34 | 72.38 | 58.52 | 68.82 | 58.33 | 68.6 | **72.56** | **81.86** |
| 3 | 96.5 | 69.41 | 81.24 | 68.86 | 80.91 | 62.03 | 75.27 | 61.52 | 74.88 | 68.81 | 79.63 | 67.3 | 78.68 | **76.62** | **86.21** |
| Average | 96.5 | 64.05 | 74.76 | 64.1 | 74.79 | 63.59 | 75.1 | 63.87 | 75.22 | 66.74 | 77.15 | 65.96 | 76.64 | **76.96** | **85.92** |
| Thoracic spine detection | | | | | | | | | | | | | | | |
| 1 | 100 | 67.33 | 80.14 | 67.98 | 80.6 | 66.57 | 79.58 | 66.84 | 79.73 | 66.61 | 79.71 | 66.76 | 79.84 | **73.03** | **84.23** |
| 2 | 100 | 68.23 | 80.65 | 68.41 | 80.69 | 68.07 | 80.73 | 68.16 | 80.71 | 70.88 | 82.32 | 70.46 | 82.02 | **75.54** | **85.56** |
| 3 | 100 | 64.78 | 77.98 | 63.73 | 77.28 | 66.02 | 79.04 | 66.02 | 79.03 | 67.09 | 79.69 | 66.01 | 78.86 | **74.61** | **85.21** |
| Average | 100 | 66.78 | 79.59 | 66.71 | 79.52 | 66.89 | 79.78 | 67.01 | 79.82 | 68.19 | 80.57 | 67.74 | 80.24 | **74.39** | **85** |

Table 4.3: Performance of Localization of Lumbar and Thoracic Spine regions

20% of each coordinate size. The experiments were repeated three times, each time splitting the dataset into 105 for training and 27 for testing for the lumbar spine, and 115 for training and 29 for testing for the thoracic spine. The results are shown in Table 4.3.The last column is the mean of the best bounding box predicted by the three models.

# 6 Conclusion

We have presented a novel method of 3D localization that combines Deep Reinforcement Learning with Imitation Learning. Localization helps to narrow down the focus and facilitate further analysis of the ROI. The method was applied to localization of vertebrae regions in 3D CT datasets, however it can be applied to any ROI in image datasets as the algorithm makes no assumptions on the dataset. It is important to note that the number of variations in the datasets is potentially huge. With a limited training set, the results are quite promising, with best average Jaccard Index/Dice Coefficient of 76.96%/85.92% for Lumbar spine and 74.39%/85% for Thoracic spine.

# References

[1] Zhigang, P., et al. *Automated Vertebra Detection and Segmentation from the Whole Spine MR Images*in 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. 2005.

[2] Pekar, V., et al., *Automated planning of scan geometries in spine MRI scans*in Proceedings of the 10th international conference on Medical image computing and computer-assisted intervention - Volume Part I. 2007, Springer-Verlag: Brisbane, Australia. p. 601-608.

[3] Klinder, T., et al., *Automated model-based vertebra detection, identification, and segmentation in CT images*Medical Image Analysis, 2009. 13(3): p. 471-482.

[4] Steinwart, I. and A. Christmann, *Support Vector Machines, in Support Vector Machines,*2008, Springer New York: New York, NY. p. 1-20.

[5] Glocker, B., et al. *Automatic Localization and Identification of Vertebrae in Arbitrary Field-of-View CT Scans.* in Medical Image Computing and

Computer-Assisted Intervention MICCAI 2012. 2012. Berlin, Heidelberg: Springer Berlin Heidelberg.

[6] Zhan, Y., et al., *Robust MR spine detection using hierarchical learning and local articulated model.* Med Image Comput Comput Assist Interv, 2012. 15(Pt 1): p. 141-8.

[7] Lootus, M., T. Kadir, and A. Zisserman, *Vertebrae Detection and Labelling in Lumbar MR Images.* Vol. 17. 2014. 219-230.

[8] Suzani, A., et al., *Fast Automatic Vertebrae Detection and Localization in Pathological CT Scans - A Deep Learning Approach,* in Medical Image Computing and Computer-Assisted Intervention MICCAI 2015. 2015. p. 678-686.

[9] Sekuboyina, A., et al., *A Localisation-Segmentation Approach for Multi-label Annotation of Lumbar Vertebrae using Deep Nets.* 2017.

[10] Janssens, R., G. Zeng, and G. Zheng, *Fully Automatic Segmentation of Lumbar Vertebrae from CT Images using Cascaded 3D Fully Convolutional Networks* 2017.

[11] Ghesu, F.C., et al., *Multi-Scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017: p. 1-1.

[12] Li, Y., *Deep Reinforcement Learning: An Overview.* 2017: Cornell Library 2017.

[13] Caicedo, J.C. and S. Lazebnik, *Active Object Localization with Deep Reinforcement Learning.* 2015.

[14] Hussein, A., et al., *Imitation Learning: A Survey of Learning Methods.* ACM Comput. Surv., 2017. 50(2): p. 1-35.

[15] Girshick, R., et al., *Rich feature hierarchies for accurate object detection and semantic segmentation.* 2014.

[16] Girshick, R., Fast R-CNN. 2015.