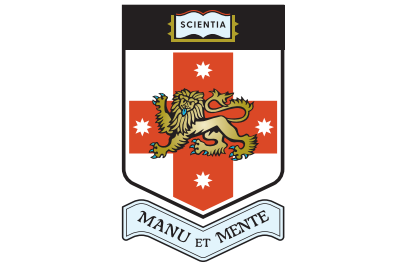
Localization of Lumbar and Thoracic Vertebrae in 3D CT Datasets by Combining Deep Reinforcement Learning with Imitation Learning

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-201803
October 2018



UNSW
SYDNEY

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

Abstract

Landmark detection and 3D localization in CT datasets is challenging due to the natural variability of human anatomical structures. We present a novel approach to lumbar and thoracic vertebrae localisation that combines 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. Two different 3D Convolutional Neural Networks (CNN) were utilized, one for learning the navigation in the coordinate directions and the other for predicting the size of the bounding box dimensions. Deep Reinforcement Learning was used to learn the direction of navigation, with random search replaced by guided search using Imitation Learning. The method achieved mean 3D Jaccard Index of 69.96% / 67.75% for lumbar spine (training on 62 datasets, testing on 20) / thoracic spine (training on 74 datasets, testing on 20).

Index Terms: 3D Localisation, Deep Reinforcement Learning, Imitation Learning, Convolutional Neural Networks, Intersection Over Union, Jaccard Index

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 using 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. The rest of the paper is organized as follows: Related topics, background and proposed work are in sections 2 and 3, the method in section 4, experiments and results in section 5 and conclusion in section 6.

2 Related Work

Traditional methods for vertebrae detection require prior knowledge vertebrae locations usually obtained from manual identification or statistical modelling and detectors based on Geometric structures [1] [2] [3] [4] and the Generalized Hough Transform [5] have been used. Machine learning methods have also been employed along with feature descriptors: Support Vector Machines[6] [7] Regression Trees [8], Regression Forests [9] [10], Adaboost [11], Marginal Space Learning [12] [13] and Deformable Parts Model [14]. Many methods require a priori knowledge of vertebrae visibility and are therefore difficult to evaluate. It is also difficult to compare methods as they were evaluated on different datasets. 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 [15] [16], joint CNN model [17], Multi-layered Perceptron (MLP) [18] [19] and 3D CNN [20]. This paper proposes a method based on Deep Reinforcement learning and Imitation learning to achieve 3D localisation of vertebrae in human 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 is an area that has seen major successes in recent times [21] combining the representation power of CNNs with Reinforcement learning. Using 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

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

The goal is to maximize this reward. The expected future discounted rewards for a given action a in a state s for a policy π 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:

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

where α is the learning rate. The agent has two choices in a state:

i explore by selecting a random action with probability ϵ

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 [22] [23]. However, bounding box localization in 3D has remained a challenge due to high computation resource requirements. Recently Deep Reinforcement Learning has been used for detection of anatomical landmarks in 3D CT datasets [24] [25] 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. Although it can be applied to such situations, reinforcement learning assumes no prior knowledge. Learning is achieved by performing random searches and the technique is more appropriate for gaming applications to determine strategies for navigation. For landmark detection, it may be more relevant and less complex for the agent to be trained in a guided manner. A navigation strategy to locate a landmark is illustrated in Figure 3.1. A simple strategy of navigating in the coordinate direction that is at maximum distance from the current location to the centre of the ground truth should suffice. We posit that it may be more 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 [26]. Unlike Reinforcement learning where the task of associating a state to actions is learned over several iterations, Imitation learning associates states with actions chosen by the expert. This converts the task to one of supervised learning of the mapping from states to expert actions.

There are variations in the literature in the interpretation of the term localization to identify an ROI. Some methods involve locating the centre of the ROI. Another approach is to surround the ROI with a bounding box (2D or 3D depending on the image or dataset). In this work localization will mean the

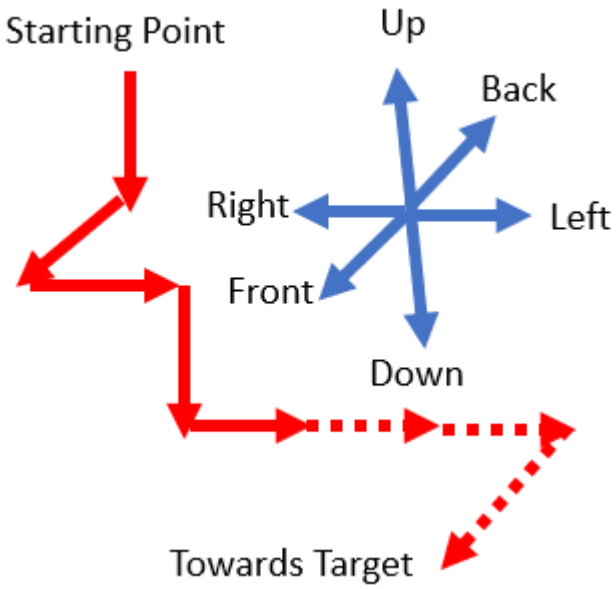


Figure 3.1: Red arrows show a navigation trajectory. Blue arrows show possible directions at each state

second interpretation and involves identifying a 3D bounding box around the lumbar/thoracic vertebrae.

Our approach to 3D bounding box localization combines the Deep Q learning algorithm [21] with Imitation learning when searching for an ROI from a predefined starting point in the image.

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 are being 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. So far 82 abdominal and 94 chest 3D datasets have been collected. The data set is manually annotated and verified by the radiologist to identify the two diagonally opposite corner points of a 3D bounding box around the ROI. 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:

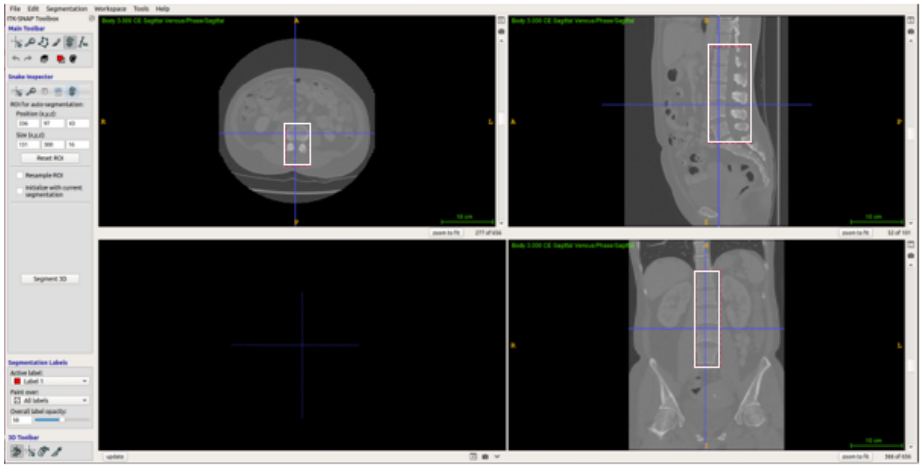


Figure 4.1: 3D Bounding Box annotation using ITK-SNAP shown by white boxes in 3 planes.

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

---

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
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 2 Convolution layers is 5x5x5 while that of the third layer is 3x3x3. 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.

We use Intersection over Union (IOU) of the predicted bounding box with the ground truth box for evaluating the localization. We use standard 50% threshold level for IOU for detection as used in ImageNet and Regions with

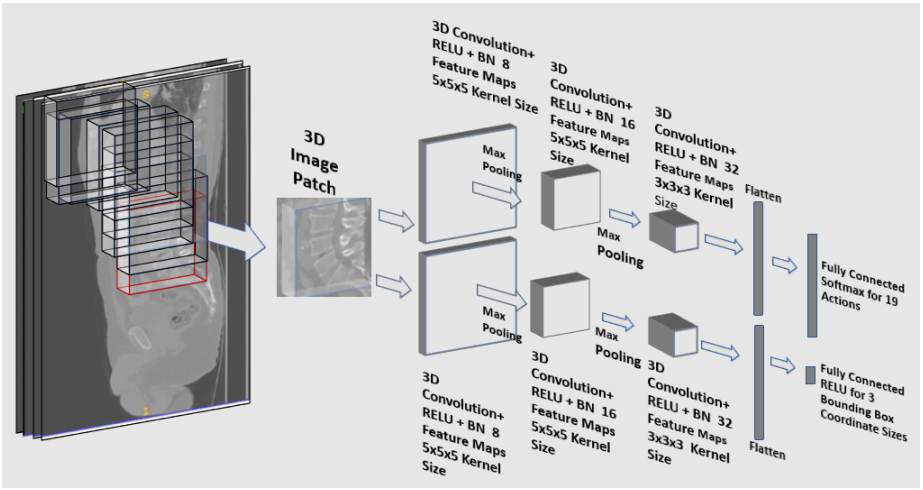


Figure 4.2: Navigation of the bounding box to the Region of Interest (ROI). The red bounding box is the target ROI

CNN for 2D bounding boxes [27-29]. 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 of 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, the 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 2 Convolution layers is 5x5x5 while that of the third is 3x3x3.

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.

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 three 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

iii Ensemble of the bounding box with dimensions using the annotations from the training set, along with the predicted bounding boxes from i and ii using the mean, maximum and minimum values of the sizes respectively

		Bounding Box Predicted by the Teminating State		Mean of the Predicted bounding Boxes of the last 10 states		Ensemble Mean*		Ensemble Maximum*		Ensemble Minimum*	
Model	Detection %	Jaccard Index(IOU)	Dice Coefficient	Jaccard Index(IOU)	Dice Coefficient	Jaccard Index(IOU)	Dice Coefficient	Jaccard Index(IOU)	Dice Coefficient	Jaccard Index(IOU)	Dice Coefficient
Lumbar1	**100**	71.86	83.35	71.97	83.27	71.69	83.08	73.51	84.37	69.50	81.71
Lumbar2	**95**	66.75	77.72	66.75	77.55	66.41	77.44	67.40	78.12	65.00	76.39
Lumbar3	**95**	70.59	82.07	71.18	82.78	70.40	82.06	67.81	80.25	69.82	81.50
Average	**96.67**	**69.73**	**81.05**	**69.96**	**81.20**	**69.50**	**80.86**	**69.57**	**80.91**	**68.11**	**79.87**

Table 4.3: showing results of lumbar spine localization after training with 62 datasets on a test set of 20.
*Ensemble of the bounding box predicted from last state, predicted mean of the last 10 states and mean of coordinate sizes from the training set

		Bounding Box Predicted by the Teminating State		Mean of the Predicted bounding Boxes of the last 10 states		Ensemble Mean *		Ensemble Maximum *		Ensemble Minimum *	
Model	Detection %	Jaccard Index(IOU)	Dice Coefficient	Jaccard Index(IOU)	Dice Coefficient	Jaccard Index(IOU)	Dice Coefficient	Jaccard Index(IOU)	Dice Coefficient	Jaccard Index(IOU)	Dice Coefficient
Thoracic1	**100**	70.63	82.36	69.61	81.58	70.04	82.02	69.09	81.41	70.16	81.96
Thoracic2	**100**	66.37	79.41	67.43	80.10	66.85	79.69	64.50	77.88	65.56	78.61
Thoracic3	**100**	65.55	78.92	64.75	78.27	66.36	79.47	63.52	77.31	63.95	77.69
Average	**100.00**	**67.52**	**80.23**	**67.26**	**79.99**	**67.75**	**80.39**	**65.70**	**78.86**	**66.55**	**79.42**

Table 4.4: Results of thoracic spine localization after training with 74 datasets on a test set of 20
*Ensemble as in Table 4.3

5 Experiments and Results

The training was run for about 25 episodes on a Keras/Tensorflow platform. The learning rate was set to 0.00001. The starting point for navigation was set around 20% of each coordinate size. The experiments were repeated three times, each time splitting the data set into 62 for training and 20 for testing for the lumbar spine and 74 for training and 20 for testing for thoracic spine. The results are shown in Table 4.3 and Table 4.4 respectively. We achieved an average detection rate of 96.67% for the lumbar spine and 100% for the thoracic spine. There are not many significant differences between the performance of the methods in the three repetitive experiments. However, there are differences in the averages of the individual cases and repetitive results.

6 Conclusion

We have presented a novel method of 3D localization combining 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. It is important to note that the number of variations in the datasets are potentially huge. With a limited training set, the results are promising even though the volumetric measures (Dice, IOU) merit more improvement. The models are being continuously trained as more data becomes available and as more variations are learned the performance is expected to further improve on those fronts.

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] Darko, ., et al., Automated detection of spinal centrelines, vertebral bodies and intervertebral discs in CT and MR images of lumbar spine Physics in Medicine Biology, 2010. 55(1): p. 247.

[4] Darko, ., et al., Parametric modelling and segmentation of vertebral bodies in 3D CT and MR spine images. Physics in Medicine Biology, 2011. 56(23): p. 7505.

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

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

[7] Oktay, A.B. and Y.S. Akgul. *Localization of the Lumbar Discs Using Machine Learning and Exact Probabilistic Inference.* in Medical Image Computing and Computer-Assisted Intervention MICCAI 2011. 2011. Berlin, Heidelberg: Springer Berlin Heidelberg.

[8] Schmidt, S., et al. *Spine Detection and Labeling Using a Parts-Based Graphical Model.* in Information Processing in Medical Imaging. 2007. Berlin, Heidelberg: Springer Berlin Heidelberg.

[9] 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.

[10] Chengwen, C., et al., *Fully Automatic Localization and Segmentation of 3D Vertebral Bodies from CT/MR Images via a Learning-Based Method.* PLoS ONE, 2015. 10(11): p. e0143327.

[11] 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.

[12] Kelm, B.M., et al. *Detection of 3D Spinal Geometry Using Iterated Marginal Space Learning.* in Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging. 2011. Berlin, Heidelberg: Springer Berlin Heidelberg.

[13] Michael Kelm, B., et al., *Spine detection in CT and MR using iterated marginal space learning.* Medical Image Analysis, 2013. 17(8): p. 1283-1292.

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

[15] 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.

[16] Suzani, A., et al., *Deep learning for automatic localization, identification, and segmentation of vertebral bodies* in volumetric MR images, in Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling. 2015.

[17] Chen, H., et al., *Automatic Localization and Identification of Vertebrae in Spine CT via a Joint Learning Model with Deep Neural Networks,* in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. 2015. p. 515-522.

[18] Shen, W., et al., *Automatic localization of vertebrae based on convolutional neural networks,* in Medical Imaging 2015: Image Processing. 2015.

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

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

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

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

[23] Kong, X., et al., *Collaborative Deep Reinforcement Learning for Joint Object Search.* 2017.

[24] Ghesu, F.C., et al., *Medical Image Computing and Computer-Assisted Intervention* – MICCAI 2016. Lecture Notes in Computer Science. 2016.

[25] 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.

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

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

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

[29] Ren, S., et al., *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.* 2016.