Prostate Size Inference from Abdominal Ultrasound Images with Patch Based Prior Information
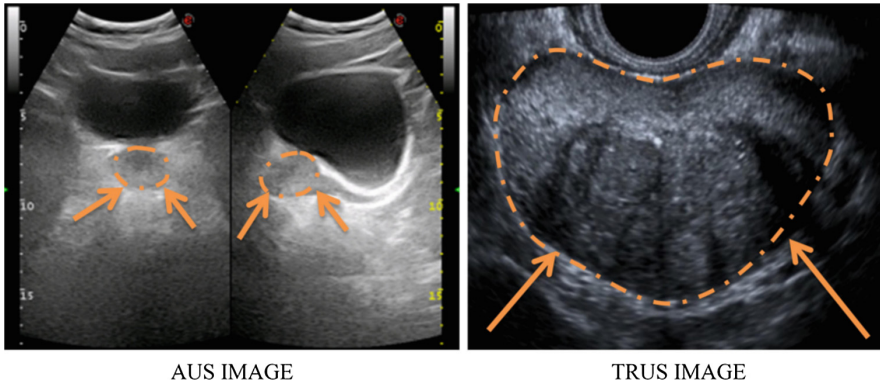
Nur Banu Albayrak, Emrah Yildirim, and Yusuf Sinan Akgul[✉]

Department of Computer Engineering, Gebze Technical University, Kocaeli, Turkey
{albayrak,yildirim,akgul}@gtu.edu.tr

Abstract. Prostate size inference from abdominal ultrasound images is crucial for many medical applications but it remains a challenging task due to very weak prostate borders and high image noise. This paper presents a novel method that enforces image patch prior information on multi-task deep learning followed by a global prostate shape estimation. The patch prior information is learned by multi-task Deep Convolutional Neural Networks (DCNNs) trained on multi-scale image patches to capture both local and global image information. We produce tens of thousands of image patches for the DCNN training that needs a large amount of training data which usually is not available for medical images. The three learned tasks for the DCNN are the distance between the patch center and the nearest contour point, the angle of the line segment between the patch center and the prostate center, and the contour curvature value for the patch center. During the prostate shape inference time, the labels returned from the multi-task DCNN are used in a global shape fitting process to obtain the final prostate contours which are then used for size inference. We performed experiments on transverse abdominal ultrasound images which are very challenging for automatic analysis.

1 Introduction

Measuring prostate size is crucial for many medical applications. As a cheap, portable, harmless, and real-time technology, ultrasound (US) is one of the most efficient technologies among many other medical imaging technologies like Magnetic Resonance (MR) and Computed Tomography (CT). Trans-rectal ultrasound (TRUS) and abdominal ultrasound (AUS) are frequently used US technologies for imaging prostate in various diagnostic and therapeutic applications. TRUS technology supplies a better imaging quality than the AUS technology with a higher signal to noise ratio (SNR) and a bigger view of prostate with no other anatomic structures. Despite its imaging quality advantages, TRUS technology is difficult to use regularly during successive radiotherapy sequences [2] due to patient discomfort [5]. AUS is a frequently applied alternative technique where TRUS is not applicable. Figure 1 shows samples of AUS and TRUS images.

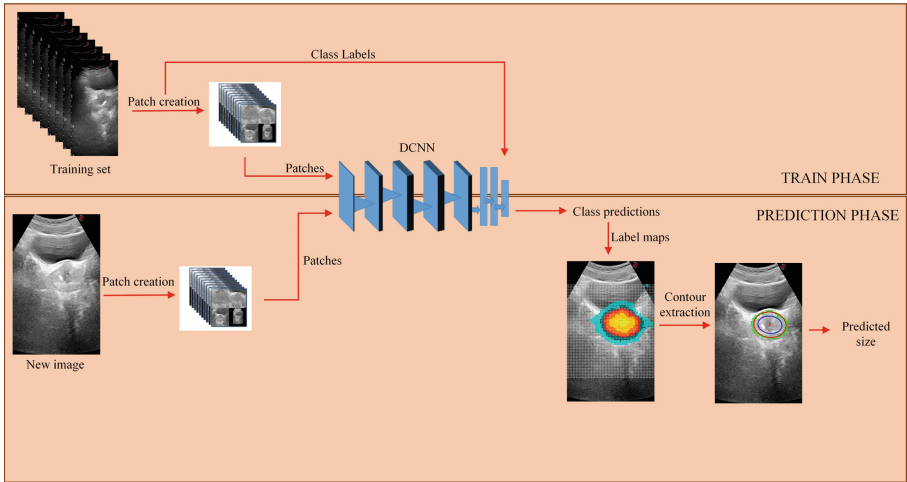<center>AUS IMAGE   TRUS IMAGE</center>

Fig. 1. Samples of transverse (left) and sagittal (middle) AUS images, and a TRUS image (right). The arrows and contours show the respective prostate borders.

Generally, clinical experts measure prostate size from US images manually in real time. However, both intra and inter observer variability is high on US images due to factors caused by imaging quality, personal experience, and human error [4]. Automated prostate size inference systems are required to overcome these limitations. However, this task is very challenging for computers. Automatic contour extraction and segmentation of prostate has become a popular research area to infer the size of the prostate by analysing US images. Although there are many studies for the analysis of the prostate in TRUS images [8,14,15], there are only a few reported AUS analysis work in the literature. [7] uses an image filtering and prior shape based approach to find the prostate border in AUS images. However, this method requires an expert initialization near the target prostate region. We argue that a fully automatic method is required to have robust systems reducing the influence of personal experience and human error.

To build a fully automatic prostate size inference system, we need a method which has two key features: First, the method should work with raw images by learning both object-level and lower-level image features. Second, the method should detect both prostate location by learning global information and prostate boundary by learning local information.

The recent researches show that Deep Convolutional Neural Networks (DCNNs) have a great potential of learning both the intermediate features and final task simultaneously from raw images. Different levels of abstract information from any type of images can be captured using DCNNs [11]. Lower levels of the network would capture features like edges, while the upper levels would capture object or class level features. These features of DCNNs correspond to the first required key feature of a fully automatic system as mentioned above. The foremost difficulty in using DCNNs in AUS image analysis is that they require immense amount of training data to extract useful features for a complex task like contour extraction or segmentation of prostate. The huge number of parameters to be learned can easily result in over-fitting if sufficient data is not fed to the network during training. This problem raises especially in domains where

Fig. 2. The general architecture of the system.

collecting data or hand labeling it is either very expensive or difficult. Medical imaging is one of these domains [3]. Most ultrasound image sets contain only at most a few hundred labeled medical images which is too small compared to more general sets with very large number of hand labeled images such as ImageNet dataset which contains more that 12 million images and 22 K classes [6]. There are methods to address the data scarcity in deep learning. One of them is pre-training with unsupervised data to learn generic features and then fine tuning the network for the task at hand in a supervised manner. However, there is no guarantee that pre-training will not cause suboptimal models and perform poorly than a fully supervised training [1]. Thus, deep learning approaches in medical image analysis were used mainly as feature extraction tools [12] or as unsupervised learning tools [13].

To take advantages of DCNNs while overcoming the dataset size limitation problem, we propose to train DCNNs with the image patch prior information to extract 2D closed contour of the prostate which is then used for size inference. We have collected a dataset of AUS images from 210 samples. Since the dataset is too small for a general DCNN training, we employ multi-task DCNNs trained with small image patches for obtaining lower-level abstract information as contour-closeness maps and then use contour fitting methods on these maps to recover the 2D closed contour, which can be considered as the object-level information.

The main contribution of this paper is to create the image patches by concatenating sub-patches of different scales and classify them in a multi-task system to combine the local and global image information which corresponds to the second required key feature of the fully automatic system mentioned above. Tens of thousands of such patches with automatically produced label data can be obtained from our small dataset, which is sufficient to train a DCNN with many layers. Three different tasks are chosen to determine the labels of the multi-task
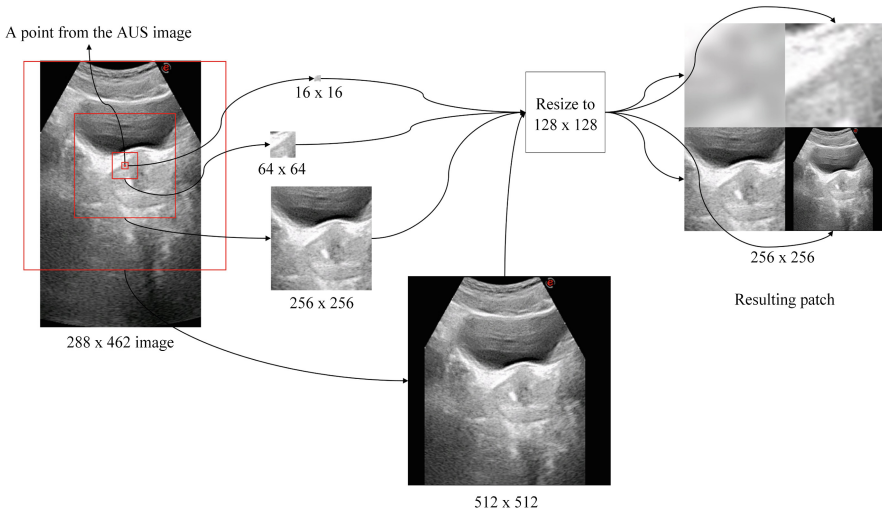
learning. These are the closeness of the patch center to the nearest contour point, the angle of the line segment between the patch center and prostate center, and the curvature of the contour line where the patch center is located. During the shape inference time, patches extracted from a test image in a sliding-window manner. Each patch is fed to the network to obtain the estimated class labels which produce a map of contour-closeness. By fitting a novel ellipse model to the map, we obtain the final contours.

The rest of this paper is organized as follows: In Sect. 2, we define our patch extraction process. In Sects. 3 and 4, we explain the training and prediction phases of our method, respectively. In Sect. 5, we present our experiments and show our quantitative results. Section 6 provides concluding remarks about this work.

2   Patch Extraction Process

Using patches composed of sub-patches at different scales prevents the loss of global information caused by reducing the representation from whole image level to patch level, while providing a patch prior information enriched by both global and local image information. To take advantage of this structure, we use patches composed of four sub-patches of four different scales with convergent centers. We extract sub-patches of 16×16, 64×64, 256×256, and 512×512 pixels and our resulting patch is 256×256 pixels.

Four equisized patches of different scales are extracted centering a given point and the patch for the point is created by concatenating these sub-patches. Figure 3 shows the patch extraction process at a given point $I_j(x, y)$
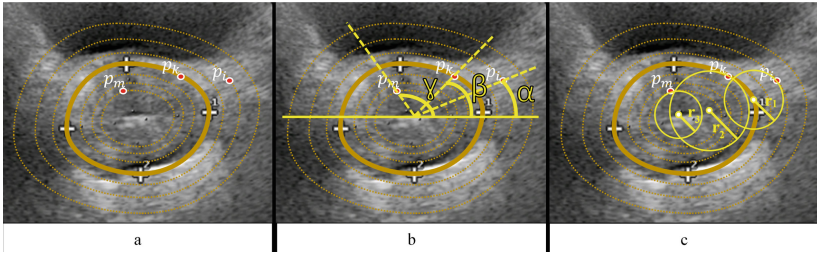

Fig. 3. Patch extraction process at a given point of image I_j. The process extracts four sub-patches of different scales. Then the final patch is created by resizing and concatenating these sub-patches.
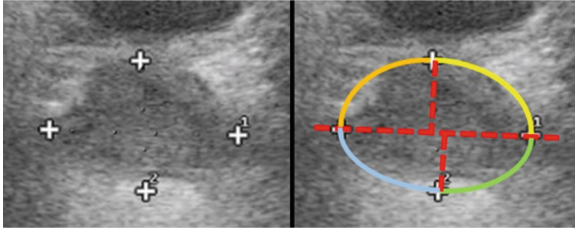
of image I_j. First, four sub-patches of different scales with convergent centers are extracted from the image, $s_1(I_j(x,y))$, $s_2(I_j(x,y))$, $s_3(I_j(x,y))$, $s_4(I_j(x,y))$. Then these sub-patches are resized and concatenated to obtain the final patch $P(I_k^{c_1,c_2,c_3}(x,y))$, where c_1, c_2, and c_3 are the class labels of the three tasks in case of the training phase.

3 Training Phase

Three different tasks are chosen for the multi-task learning process. The labels of the first task, c_1 represents the distance of the patch center to the closest contour point. We create several concentric image layers by scaling up or down the contour layer to quantize the distance values into the number of classes. As shown in Fig. 4a, class labels for the first task of three sample points, $p_i^{c_1}$, $p_k^{c_1}$, and $p_m^{c_1}$ are determined according to the layers they are on. Assuming that the class label for the innermost layer is 1, the respective values for $p_i^{c_1}$, $p_k^{c_1}$, and $p_m^{c_1}$ will be 6, 4, and 1. Given P as the line segment between the center of the prostate and patch center p_i, the labels of the second task, c_2 represents the angle between the line segment P and the x-axis. Figure 4b shows the calculation of the angle values of three sample points for the second task. After the calculation, each angle value is quantized into the number of classes determined for the second task, which is 6 for our experiments. The labels of the third task, c_3 represents the curvature value for the patch center which is calculated considering the neighbouring points on the same layer. Curvature for three points $p_{(i-1)}$, p_i, and $p_{(i+1)}$ is the radius of the circle which the three points are on. Figure 4c shows the calculation of the curvature value for three points. Curvature value is also quantized into the number of classes after calculation.

Fig. 4. Determination of class labels for the three learning tasks. Solid contour shows the contour layer and dotted lines show the other layers. p_i, p_k, and p_m are three sample points from outer layers, contour boundary and inner layers. In (a), class labels of the points for the first learning task, c_1 is determined considering the layer it is on. In (b), the class labels for the second learning task of the sample points, c_2 is determined by quantizing α, β, and γ into the number of angle classes. In (c), the class labels for the third learning task of the sample points, c_3 is determined by quantizing r_1, r_2, and r_3 into the number of curvature classes. (Color figure online)

Fig. 5. Use of four ellipse sections for the train data contour creation. Right: expert contours obtained by using the four points and four ellipse sections. Left: four points of expert.
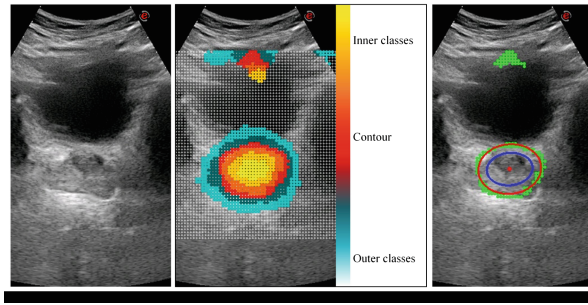
Since our dataset has only four landmark annotations as shown on the left column of Fig. 5, we obtain complete training contours from the four landmarks by fitting four different ellipse sections as shown on the right column of Fig. 5. The horizontal line between two landmarks which shows the prostate width is considered as the intersection of the horizontal axes of four ellipses. The vertical axes of the four ellipses are the normal to the horizontal axes from the two landmarks.

We train a multi-task DCNN with the train patches which are created according to the patch creation method explained in Sect. 2. Training phase procedures are shown in training phase of Fig. 2.

4   Prediction Phase

Given a new image we remove the 1/4 of the height of the image from the top and bottom rows. Then we extract patches using the method explained in Sect. 2, by a sliding window approach. After patch extraction process, we use the previously trained multi-task DCNN to obtain class predictions for each patch. This way we obtain a contour-closeness map for each test image. These procedures are shown in prediction phase of Fig. 2.
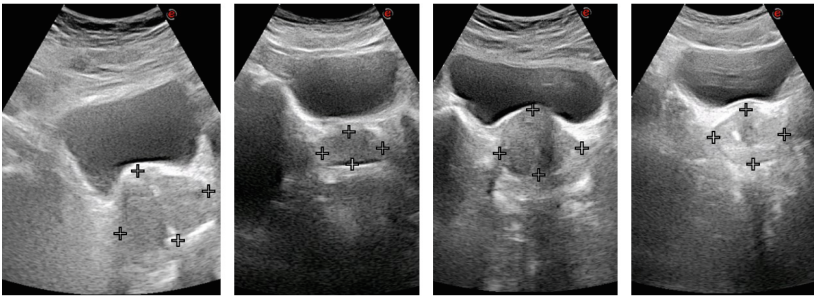
We can find the best ellipse that fits to a group of points. To employ this for the contour extraction process, we create a candidate group of points according to the map values of the test points. First, we use the points which are classified as inside the prostate to find a good start location to estimate the contour. For this purpose, we use the three layers' test points to create a new map. Then we find the largest blob which is supposed to show the real location of the prostate. An ellipse is fitted to the blob and enlarged until enough number of points which are detected as on the contour are inside the ellipse. We determine the number of points by checking if at least one point is inside the ellipse for its each quarter. The axes of the resulting ellipse are the width and height of our predicted prostate. Figure 6 shows the contour extraction process.

Fig. 6. Extraction of contour from a class prediction map. From left to right: first, a new image. Second, class predictions map. Third, the class predictions of being inside the prostate. Fourth, blobs detected. Fifth, the first ellipse (blue) which is fit to the largest blob, and the second ellipse (red) which is enlarged according to the points (green) classified as on the contour. (Color figure online)

5 Experiments

We use transverse abdominal ultrasound images to demonstrate the performance of our method on 2D size inference. 210 images are used for experiments. Figure 7 shows some sample images from our dataset. We applied 5-fold cross validation, resulting with 180 train and 30 test images in each fold. That is 150 of the images are used for testing. Train patches with labels are created with patch creation and label determination methods explained in Sects. 2 and 3.

Fig. 7. Some sample images from our dataset with expert marks.

We use Caffe [9] implementation of AlexNet [10] and use Stochastic Gradient Descent for DCNN optimization. We perform our experiments in three different cases; using only distance labels in single-task learning, using distance and angle labels in multi-task learning and using distance, angle and curvature labels in multi-task learning. We modify AlexNet to be multi-task DCNN for the multi-task learning experiments.

Table 1. Mean of size differences between expert contours and detected contours in cm^2 within contour detected results.

Fold no	Test case 1 distance (undetected contours)	Test case 2 distance + angle + (undetected contours)	Test case 3 distance + angle + curvature (undetected contours)
1	3.10 (2)	2.98	2.94 (1)
2	2.94	2.76	3.00
3	3.02 (1)	2.78	3.19
4	2.95 (1)	3.18 (1)	3.09 (2)
5	1.81 (2)	1.80	1.93
Mean	2.76	2.70	2.83

We create 7 layers including the contour layer for the first learning task to obtain 7 classes which represent the signed distance value of a patch center to the closest contour point. For the second learning task, we choose angle ranges of 60° for each class resulting with 6 classes of angles. 11 classes are obtained for the third learning task which represents the curvature value of the layer where a patch center is on.
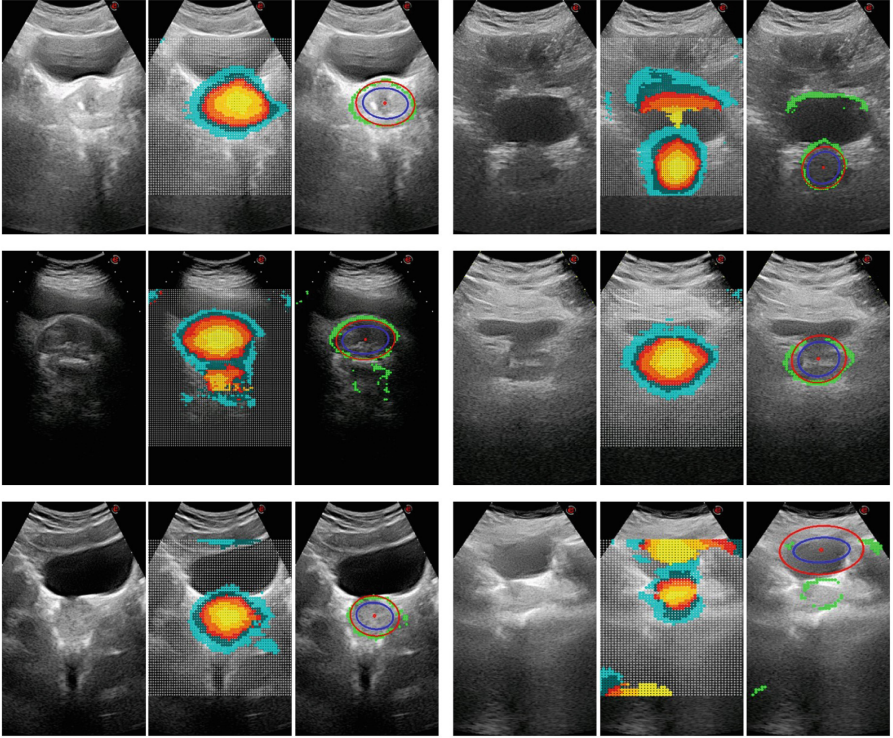
After obtaining the contour-closeness maps which are composed of distance class predictions, a final ellipse is created to predict the size of the prostate as a 2D area. The size of the prostate in 2D is calculated with (1).

$$size = \frac{width * height * \pi}{4} \qquad (1)$$

Our test results are shown in Table 1 as errors in cm^2. Each column shows a test case while each row shows a cross validation fold. There are 10 miss-detection cases which are excluded from the error results. The best DCNN architecture is the multi-task network that assigns labels based on patch distance and angle. This network has only one miss-detection out of 150 cases which is shown Fig. 8.

The literature reports inter-observer differences of dimensions measurement in mm between experienced observers as 1.5 ± 1.3 for width and 1.8 ± 2.5 for height, between experienced and less experienced observers as 1.6 ± 1.4 for width and 2.8 ± 2.4 for height on TRUS images [4]. Although we did not use the same data set and our images are of type AUS, differences of dimensions measurement in mm between less experienced observers and our results are 6.1 ± 5.3 for width and 4.9 ± 3.9 for height. The comparison results are shown in Table 2. Note that the prostate region is several times larger on TRUS images than AUS images (see Fig. 1). As a result, it is experienced that a single pixel error on an AUS image would translate to several times larger errors in mm. Considering the other problems with the AUS images that were listed in Sect. 1, we argue that our error results on AUS images are comparable to errors on TRUS images.

Fig. 8. Samples from our results. First and fourth columns show new test images. Second and fifth columns show class prediction maps. Third and sixth columns show detected contours. Colour scale for the class predictions maps are given in Fig. 6. The only misdetection in our best test case (distance + angle multi-task learning case) is shown in the third row and sixth column.

Table 2. Comparison of our dimensions measurement results with [4] in mm. First column and second column shows the variations between experienced observers and experienced and less experienced observers. Third column shows variations between less experienced observers and our results.

Measurement of	exp. vs. exp. [4] (mean $\pm$ SD)	lexp. vs. exp. [4] (mean $\pm$ SD)	lexp. vs. ours. (mean $\pm$ SD)
Width (mm)	1.5 $\pm$ 1.3	1.6 $\pm$ 1.4	6.1 $\pm$ 5.3
Height (mm)	1.8 $\pm$ 2.5	2.8 $\pm$ 2.4	4.9 $\pm$ 3.9

6   Conclusions

Prostate size measurement from AUS images has many advantages. However, problems unique to AUS modality makes this task challenging for an automated analysis. We have presented a novel prostate size inference method that learns image patch level prior information that include both local and global

information about the patch. By transforming the problem input from complete images to small patches, we effectively address the data scarcity issue that usually is a problem for medical applications when used with DCNNs. We have developed a multi-scale approach while forming the patches to include both local and global information about the patches.

The proposed method is applied on transverse abdominal ultrasound images with DCNN architectures that assign more than one class label at the same time, which improves the overall performance. The complete results are available[1]. The results show that our size inference results are comparable to manual expert contours. For the future work, we plan to estimate the sagittal sizes to produce a complete 3D size measurement tool from AUS images. We also plan to make direct comparisons against inter-observer variability on our data set.

References

1. Astudillo, R.F., Amir, S., Ling, W., Silva, M., Trancoso, I.: Learning word representations from scarce and noisy data with embedding subspaces. In: ACL, vol. 1, pp. 1074–1084 (2015)
2. Betrouni, N., Vermandel, M., Pasquier, D., Maouche, S., Rousseau, J.: Segmentation of abdominal ultrasound images of the prostate using a priori information and an adapted noise filter. Comput. Med. Imaging Graph. **29**(1), 43–51 (2005)
3. Carneiro, G., Nascimento, J.C., Freitas, A.: The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. IEEE Trans. Image Process. **21**(3), 968–982 (2012)
4. Choi, Y.J., Kim, J.K., Kim, H.J., Cho, K.S.: Interobserver variability of transrectal ultrasound for prostate volume measurement according to volume and observer experience. Am. J. Roentgenol. **192**(2), 444–449 (2009)
5. De Sio, M., Darmiento, M., Di Lorenzo, G., Damiano, R., Perdonà, S., De Placido, S., Autorino, R.: The need to reduce patient discomfort during transrectal ultrasonography-guided prostate biopsy: what do we know? BJU Int. **96**(7), 977–983 (2005)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition 2009, CVPR 2009, pp. 248–255. IEEE (2009)
7. Ghanei, A., Soltanian-Zadeh, H., Ratkewicz, A., Yin, F.F.: A three-dimensional deformable model for segmentation of human prostate from ultrasound images. Med. Phys. **28**(10), 2147–2153 (2001)
8. Ghose, S., Oliver, A., Martí, R., Lladó, X., Vilanova, J.C., Freixenet, J., Mitra, J., Sidibé, D., Meriaudeau, F.: A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. Comput. Methods Programs Biomed. **108**(1), 262–287 (2012)
9. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678. ACM (2014)

---

[1] http://vision.gyte.edu.tr/pages/projects.html.

10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

11. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 609–616. ACM (2009)

12. Liao, S., Gao, Y., Oto, A., Shen, D.: Representation learning: a unified deep learning framework for automatic prostate MR segmentation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 254–261. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40763-5_32

13. Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., Eric, I., Chang, C.: Deep learning of feature representation with multiple instance learning for medical image analysis. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1626–1630. IEEE (2014)

14. Yuan, J., Qiu, W., Ukwatta, E., Rajchl, M., Tai, X.C., Fenster, A.: Efficient 3D endfiring trus prostate segmentation with globally optimized rotational symmetry. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2211–2218 (2013)

15. Zhang, Y., Matuszewski, B.J., Histace, A., Precioso, F., Kilgallon, J., Moore, C.: Boundary delineation in prostate imaging using active contour segmentation method with interactively defined object regions. In: Madabhushi, A., Dowling, J., Yan, P., Fenster, A., Abolmaesumi, P., Hata, N. (eds.) Prostate Cancer Imaging 2010. LNCS, vol. 6367, pp. 131–142. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15989-3_15