Deep Adversarial Context-Aware Landmark Detection for Ultrasound Imaging

Ahmet Tuysuzoglu[1(✉)], Jeremy Tan[2], Kareem Eissa[1], Atilla P. Kiraly[1], Mamadou Diallo[1], and Ali Kamen[1]

[1] Siemens Healthineers, Medical Imaging Technologies, Princeton, NJ, USA
ahmet.tuysuzoglu@siemens-healthineers.com
[2] Imperial College London, London, UK

Abstract. Real-time prostate gland localization in trans-rectal ultrasound images is required for automated ultrasound guided prostate biopsy procedures. We propose a new deep learning based approach aimed at localizing several prostate landmarks efficiently and robustly. Our multitask learning approach primarily makes the overall algorithm more contextually aware. In this approach, we not only consider the explicit learning of landmark locations, but also build-in a mechanism to learn the contour of the prostate. This multitask learning is further coupled with an adversarial arm to promote the generation of feasible structures. We have trained this network using ∼4000 labeled trans-rectal ultrasound images and tested on an independent set of images with ground truth landmark locations. We have achieved an overall Dice score of 92.6% for the adversarially trained multitask approach, which is significantly better than the Dice score of 88.3% obtained by only learning of landmark locations. The overall mean distance error using the adversarial multitask approach has also improved by 20% while reducing the standard deviation of the error compared to learning landmark locations only. In terms of computational complexity both approaches can process the images in real-time using a standard computer with a CUDA enabled GPU.

1 Introduction

Multi-parametric MRI can greatly improve prostate cancer detection and can also lead to a more accurate biopsy verdict by highlighting areas of suspicion [1]. Unfortunately, MR-guided procedures are costly and restrictive, whereas ultrasound guidance offers more flexibility and can exploit added MR information through fusion [9]. A key step in diagnostic MR and live trans-rectal ultrasound registration is the real-time, automated prostate gland localization within the

---

A. Tuysuzoglu and J. Tan—Equal contribution.

Disclaimer: This feature is based on research, and is not commercially available. Due to regulatory reasons its future availability cannot be guaranteed.

ultrasound image. This localization could be achieved by automatically identifying image landmarks on the border of the prostate. This task by itself is in general challenging due to low tissue contrast leading to fuzzy boundaries and varying prostate gland sizes in the population. Furthermore, prostate calcifications cause shadowing within the ultrasound image hindering the observation of the gland boundary. An example of this case is shown in Fig. 1(a). Learning these landmark locations is further complicated by inherent label noise as these landmarks are not defined with absolute certainty. A small inter-slice variability in prostate shape could result in rather larger deviation in the landmark locations, which are placed by expert annotators. Our analysis of this uncertainty is further explained in Sect. 2.

Through an initial set of experiments we observed that individual landmark detection/regression does not yield satisfactory results as the global context in terms of how the landmarks are connected is not properly utilized. Even for expert annotators, context is essential to place the challenging landmarks, specifically ones in regions with little signal or cues. Incorporating topological/spatial priors into landmark detection tasks is an active area of research with broad applications. Conditional Random Fields incorporating priors have been used with deep learning to improve delineation tasks in computer vision [3,11]. In medical imaging, improving landmark and contour localization tasks through the use of novel deep learning architectures has been presented in [6,10]. In particular in [10], the authors considered the sequential detection of prostate boundary through the use of recurrent neural networks in polar coordinate transformed images; however, their method assumes that the prostate is already localized and cropped.

In this work we propose a deep adversarial multitask learning approach to address the challenges associated with robust prostate landmark localization. Our design aims to improve performance in regions, where the boundary is ambiguous, by using the spatial context to inform landmark placement. Multitask learning provides an effective way to bias a network to learn additional information that can be useful for the original task through the use of auxiliary tasks [2]. In particular, to bring in the global context, we learn to predict the complete boundary contour in addition to each landmark location to enforce the overall algorithm in being contextually aware. This multitasking network is further coupled by a discriminator network that provides feedback regarding the predicted contour feasibility. Our work shares similarities with [4], where the authors used multitasking with adversarial regularization in human pose estimation in an extensive network. Unlike the method in [4], our approach is easily trainable and can perform at high frame rates and compared to [10], it does not require prior prostate gland localization.

2   Methods

This study includes data from trans-rectal ultrasound examinations of 32 patients, resulting in 4799 images. Six landmarks distributed on the prostate

boundary are marked by expert annotators. In particular, the landmark locations are chosen to cover the anterior section of the gland (close to bladder), posterior section (close to rectum), and left and right extent of the gland considering the shape of the probe pressing into the prostate. Examples of annotations can be seen in Fig. 1(a). Nonetheless the landmarks cannot be placed with complete certainty due to poor boundaries, missing defining features, shadowing and other physiological occurrences such as calcifications. We characterized this landmark annotation uncertainty by measuring the change in landmark position in successive frames. The mean and standard deviation for each landmark position is given in Table 1. It is understood that part of this positional difference is due to probe and patient movement but nevertheless they can be treated as a lower bound for the localization error that can be achieved.

Each image is acquired as part of a 2D sweep across the prostate and all images were resampled to have a resolution of 0.169 mm/pixel and then padded or cropped so that the resulting image size is 512×512. Training data is tripled via augmentation with translation (±30–70 pixels) plus noise ($\sigma = 0.05$) and rotation (±4–7°) plus noise ($\sigma = 0.05$). We split the data into 3 sets: 23 patients for training (3717 images, 77%), 6 patients for validation (853 images, 18%), and 3 patients for testing (229 images, 5%). For all methods explained below the ultrasound data is given to the network as 2-D images.
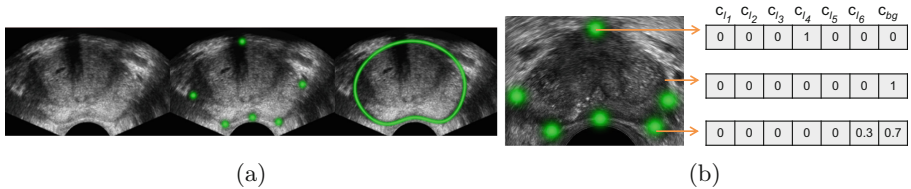
c_{l_1}	c_{l_2}	c_{l_3}	c_{l_4}	c_{l_5}	c_{l_6}	c_{bg}	
0	0	0	1	0	0	0	
0	0	0	0	0	0	0	1
0	0	0	0	0	0.3	0.7	

(a)   (b)

Fig. 1. (a) Ultrasound images with target labels: 2D Gaussian landmarks (center, green) and contours (right, green). (b) Each pixel has a distribution over 7 classes: 6 landmark classes and the background class. Moving away from the center of a landmark, the landmark probability decreases and the background probability increases.

2.1 Baseline Approach for Landmark Detection

Given the landmark locations, our approach takes a classification approach through the use of a shared background in locating the landmarks rather than the classical regression approach. The network has a 5 layer convolutional encoder and a corresponding decoder with 5×5 kernels, padding of 2, stride of 1, and a pooling factor of 2 at each layer. The number of filters in the first layer is 32; this doubles with every convolutional layer in the encoder to a maximum of 512. The decoder halves the number of filters with each convolutional layer. The final output is convolved with a 1×1 kernel into 7 channels (one for each landmark and a background class). The configuration of the convolutional, batch normalizing, rectifying, and pooling layers can be seen in Fig. 2.

We model each landmark as a 2D Gaussian function centered on the land-mark. The standard deviation of this Gaussian can in part incorporate the uncertainty involved in the landmark locations. In contrast to the regression approaches that regress locations or probability maps independently for each landmark, here we take a classification approach which couples the estimation through a shared background. For each pixel in the ultrasound image, we assign a probability distribution over 7 classes, where we treat each landmark and the background as separate classes. For a pixel that is at the center of a Gaussian for a landmark, the probability for that landmark class is 1 whereas rest of the probabilities are set to zero. These probabilities are obtained by independently normalizing each Gaussian distribution so that the maximum of the Gaussian is 1. Similarly for a pixel that does not overlap with any of the Gaussian functions, the background class has probability 1 and rest of the classes are set to zero. For a pixel that overlaps with one of the landmarks but not necessarily at the center, the probability distribution over the classes is shared between the corresponding landmark class and the background class. This is illustrated in Fig. 1(b). This framework can be trivially extended to scenarios where the Gaussian functions for the landmarks overlap. We learn a mapping of training images $\mathbf{x}$ in train-ing set $\mathbf{X}$ that represents the probability distribution of every pixel in $\mathbf{x}$ over the classes. This mapping, $S_{\mathrm{lm}}(\mathbf{x})$, is learnt through the minimization of the following supervised loss where $\mathbf{Y}_{\mathrm{lm}}$ denotes the training set labels:

$$\mathcal{L}_{\mathrm{lm}} = -\mathbb{E}_{(\mathbf{x},\mathbf{y}_{\mathrm{lm}})\sim(\mathbf{X},\mathbf{Y}_{\mathrm{lm}})}[\log S_{\mathrm{lm}}(\mathbf{x})]. \tag{1}$$

During test time the landmark locations are obtained by processing the out-put maps, i.e., by extracting the maxima. The joint prediction of landmark and background classes could help the network become more aware of the positions of each landmark relative to one another. However, this background class encom-passes the entire space wherever a landmark does not exist. As such, it does not explicitly relate the points or highlight specific image features that are relevant to the connections between points (e.g. organ contour).

2.2   Multitask Learning for Joint Landmark and Contour Detection

When deciding a landmark location, expert annotators/clinicians are equipped with the prior knowledge that the landmarks exist along the prostate boundary which is a smooth, closed contour. Motivated by this intuition we identify two distinct priors: First, the points lie along the prostate boundary, and then this boundary must form a smooth, closed contour despite occlusions. We incorpo-rate these priors through multitask learning and the use of an adversarial cost function.

In multitask learning, the network must identify a set of auxiliary labels in addition to the main labels. The main labels (in this case landmarks) help the network to learn the appearance of the landmarks; meanwhile the auxiliary labels should promote learning of complementary cues that the network may otherwise ignore. A fuzzy contour following the prostate boundary is obtained by Gaussian
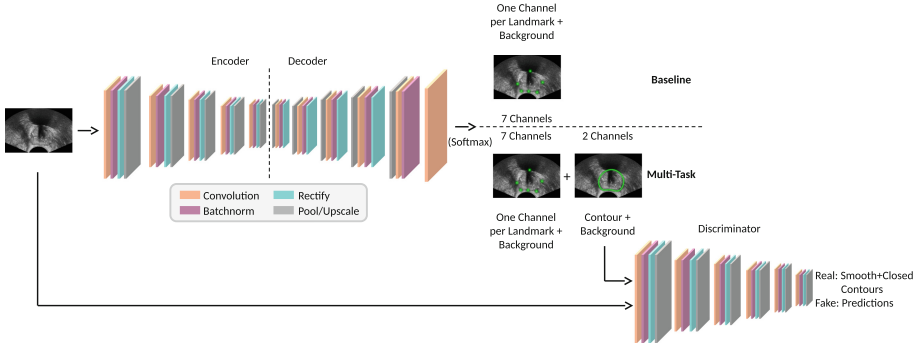
Fig. 2. Our baseline network has an encoder-decoder architecture where the receptive field size is large enough to contain the entire prostate. The multitask network outputs a boundary contour along with the landmarks which is then fed to a discriminator network to evaluate its similarity to training set samples.

blurring the spline generated by the main landmark labels. The boundary is used as an auxiliary label to incorporate the first spatial prior, that all landmarks lie on the prostate boundary. The goal of the multitask addition is to bias the network's features such that prostate boundary detection is enhanced. Since the boundary overlaps directly with the landmarks, the auxiliary task lends itself well to exploitation in the shared parameter representation. Figure 2 displays the addition of the auxiliary label for the multitask framework. Note that the network size does not increase, except for the final layer, because the parameters are shared between both tasks.

Similar to the landmark setup, we learn a mapping of training images, $S_{\mathrm{cnt}}(\mathbf{x})$, representing the likelihood of being a contour pixel by minimizing the following supervised loss, where $\mathbf{Y}_{\mathrm{cnt}}$ denotes the training set labels associated with the contour:

$$\mathcal{L}_{\mathrm{cnt}} = -\mathbb{E}_{(\mathbf{x}, \mathbf{y}_{\mathrm{cnt}}) \sim (\mathbf{X}, \mathbf{Y}_{\mathrm{cnt}})}[\log S_{\mathrm{cnt}}(\mathbf{x})]. \tag{2}$$

Discriminator Network
While the multitask framework aims to increase the network's awareness of the prostate boundary features, it does not enforce any constraint on the predicted contour shape. As such, a discriminator network is added to motivate fulfillment of the second prior, that the boundary is a smooth closed shape. This is helpful because the low tissue contrast can make it challenging for the boundary detection (learned by the multitask network) to give clean estimates without false positives. The discriminator network is trained in a conditional style where the input training image is provided together with the network generated or the real contour. The design is similar to the encoder in the main encoder-decoder network with the difference being the discriminator network is extended one layer further and the first 3 layers have a pooling factor of 4 instead of 2. These changes are made to rapidly discard high resolution details and focus the discriminator's

evaluation on the large scale appearance. We then define the discriminator loss as follows:

$$\mathcal{L}_{\mathrm{adv}_D} = -\mathbb{E}_{(\mathbf{x},\mathbf{y}_{\mathrm{cnt}})\sim(\mathbf{X},\mathbf{Y}_{\mathrm{cnt}})}[\log D\left(\mathbf{x},\mathbf{y}_{\mathrm{cnt}}\right)]$$
$$-\mathbb{E}_{(\mathbf{x}\sim\mathbf{X})}[\log\left(1 - D\left(\mathbf{x}, S_{\mathrm{cnt}}(\mathbf{x})\right)\right)]. \tag{3}$$

In [5], the authors defined the generator loss as the negative of the discriminator loss defined in Eq. 3, resulting in a min-max problem over the generator and discriminator parameters. The authors in [5] (and several others [7,8]) have also stated the difficulty with the min-max optimization problem and suggested maximizing the log probability of the discriminator being mistaken as the generator loss. This corresponds to the following adversarial loss for the landmark and contour network S:

$$\mathcal{L}_{\mathrm{adv}_S} = -\mathbb{E}_{(\mathbf{x}\sim\mathbf{X})}[\log D\left(\mathbf{x}, S_{\mathrm{cnt}}(\mathbf{x})\right)]. \tag{4}$$

Adversarial Landmark and Contour Detection Framework
The landmark and contour detection network is trained by minimizing the following functional with respect to its parameters θ_S:

$$\arg\min_{\theta_S} \ \mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{lm}} + \lambda_1\mathcal{L}_{\mathrm{cnt}} + \lambda_2\mathcal{L}_{\mathrm{adv}_S} \tag{5}$$

The discriminator is trained by minimizing $\mathcal{L}_{\mathrm{adv}_D}$ with respect to its parameters θ_D. We optimize these two losses in an alternating manner by keeping θ_S fixed in the optimization of the discriminator and θ_D fixed in the optimization of the detector network. In our experiments, we picked $\lambda_1 = 1$ and $\lambda_2 = 0.02$ using cross validation.

3   Results and Discussion

Landmark location has a range of acceptable solutions on the prostate boundary that is also visible in the noise of the annotated labels. As such, the Dice score between the spline interpolated prostate masks is used as the primary evaluation metric. In addition, the Euclidean distance between predictions and targets and the 80th percentile of this distance are calculated. Baseline Dice score and average landmark error are 88.3% and 3.56 mm respectively. The multitask approach improves these scores to 90.2% and 3.12 mm. Adversarial training further improves the results to 92.6% and 2.88 mm. In particular, note the large improvement for landmark 4 (Table 1). This is the most anterior landmark (close to bladder) which generally has the highest error due to shadowing. Also, the improvement in the standard deviation of the Dice score indicates that the adversarially regulated multitask framework produces the most robust predictions.

Figure 3 displays prediction examples given by each method. In the top row, the plain multitask approach is able to improve the right-most landmark placement, but the most anterior landmark location is still inaccurate. In such cases,

Table 1. Landmark annotation error together with error for baseline, multitask, and adversarial multitask methods in units of mm.

Metric	Noise	Baseline	Multitask	Multitask GAN
Mean landmark error $\pm$ S.D.	0.98 ± 0.28	2.11 ± 1.41	1.94 ± 1.36	1.77 ± 1.43
	1.45 ± 0.44	2.33 ± 1.28	1.90 ± 1.13	1.97 ± 0.96
	2.17 ± 0.60	4.03 ± 5.13	3.38 ± 3.68	3.41 ± 3.17
	1.99 ± 0.47	6.29 ± 6.13	6.72 ± 5.59	5.01 ± 3.90
	2.19 ± 0.74	3.44 ± 2.77	2.73 ± 1.94	3.09 ± 2.43
	1.43 ± 0.54	3.21 ± 4.05	2.02 ± 1.85	2.01 ± 1.57
Overall avg.	1.70 ± 0.51	3.56 ± 3.46	3.12 ± 2.60	2.88 ± 2.24
80th percentile	1.42	3.19	3.04	2.75
	2.05	3.44	2.85	2.72
	3.17	4.59	4.41	5.08
	2.87	8.31	9.09	7.75
	3.14	4.83	4.27	4.68
	2.03	3.71	2.75	2.90
Overall avg.	2.45	4.68	4.42	4.32
Avg. dice $\pm$ S.D.	-	$88.3\% \pm 7.3\%$	$90.2\% \pm 7.2\%$	$92.6\% \pm 3.6\%$
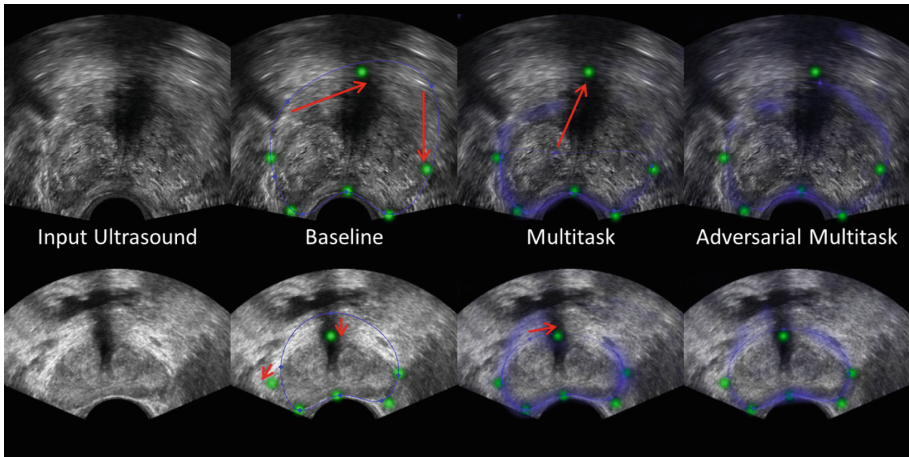
Fig. 3. Adversarially regulated multitask learning produces more complete contours resulting in better landmark placement compared to its plain counterpart. Ultrasound images with target (green) and prediction (blue diamonds, connected by spline) overlays. Red arrows indicate corrections of gross errors. Multitask predictions include an overlay of the contour prediction (blue heatmap).

features learned for boundary detection can mistakenly highlight areas with high contrast, e.g. calcification within the prostate. The adversarially trained detector improves the landmark placement significantly. In the bottom row, the boundary

prediction is also hindered by shadowing, but the proposed framework still improves the overall shape of the contour along with the landmark placements.

The multitask learning framework helps biasing the landmark placement toward the prostate boundary through shared weights of two tasks, namely landmark detection and boundary estimation. As the predicted contour is not always of high quality especially when there is signal dropouts, an adversarial regularization is used to enhance boundary estimations and subsequently provide more accurate landmark detection.

References

1. Boesen, L.: Multiparametric MRI in detection and staging of prostate cancer. Scand. J. Urol. **49**, 25–34 (2015)
2. Caruana, R.: Multitask learning. In: Thrun, S., Pratt, L. (eds.) Learning to Learn, pp. 95–113. Springer, Boston (1998). https://doi.org/10.1007/978-1-4615-5529-2_5
3. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Semantic image segmentation with deep convolutional nets and fully connected CRFs. CoRR (2014)
4. Chen, Y., Shen, C., Wei, X.S., Liu, L., Yang, J.: Adversarial posenet: a structure-aware convolutional network for human pose estimation. CoRR 2 (2017)
5. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (2014)
6. Payer, C., Štern, D., Bischof, H., Urschler, M.: Regressing heatmaps for multiple landmark localization using CNNs. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 230–238. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_27
7. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Computer Vision and Pattern Recognition (CVPR) (2017)
8. Usman, B., Saenko, K., Kulis, B.: Stable distribution alignment using the dual of the adversarial distance. arXiv preprint arXiv:1707.04046 (2017)
9. Yacoub, J.H., Verma, S., Moulton, J.S., Eggener, S., Oto, A.: Imaging-guided prostate biopsy: conventional and emerging techniques. Radiographics **32**, 819–837 (2012)
10. Yang, X., et al.: Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (2017)
11. Zheng, S., et al.: Conditional random fields as recurrent neural networks. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (2015)