Joint Learning of Image Regressor and Classifier for Deformable Segmentation of CT Pelvic Organs

Yaozong Gao[1,2], Jun Lian[3], and Dinggang Shen[1]

[1] Departments of Radiology, Chapel Hill, USA
[2] Computer Science, Chapel Hill, USA
[3] Radiation Oncology, University of North Carolina at Chapel Hill, Chapel Hill, USA

Abstract. The segmentation of pelvic organs from CT images is an essential step for prostate radiation therapy. However, due to low tissue contrast and large anatomical variations, it is still challenging to accurately segment these organs from CT images. Among various existing methods, deformable models gain popularity as it is easy to incorporate shape priors to regularize the final segmentation. Despite this advantage, the sensitivity to the initialization is often a pain for deformable models. In this paper, we propose a novel way to guide deformable segmentation, which could greatly alleviate the problem caused by poor initialization. Specifically, random forest is adopted to jointly learn image regressor and classifier for each organ. The image regressor predicts the 3D displacement from any image voxel to the organ boundary based on the local appearance of this voxel. It is used as an external force to drive each vertex of deformable model (3D mesh) towards the target organ boundary. Once the deformable model is close to the boundary, the organ likelihood map, provided by the learned classifier, is used to further refine the segmentation. In the experiments, we applied our method to segmenting prostate, bladder and rectum from planning CT images. Experimental results show that our method can achieve competitive performance over existing methods, even with very rough initialization.

1 Introduction

Prostate cancer is the second leading cause of cancer death in American men. External beam radiation therapy (EBRT) is an effective treatment option to control prostate cancer. In the planning stage of EBRT, a planning CT image is acquired. Physicians need to segment major structures such as the prostate, bladder and rectum from this CT image to design a radiation treatment plan. The efficacy of EBRT depends heavily on the quality of segmentation. As manual segmentation is time-consuming and often suffers from large inter-observer variability, it is clinically desirable to develop accurate automatic methods for segmenting CT pelvic organs.

Among various existing methods, deformable models have gained the most popularity, as it is easy to impose shape regularization. For example, Lay et al. [1] adopted the active shape model with boundary detectors to segment pelvic

N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 114–122, 2015.
DOI: 10.1007/978-3-319-24574-4_14

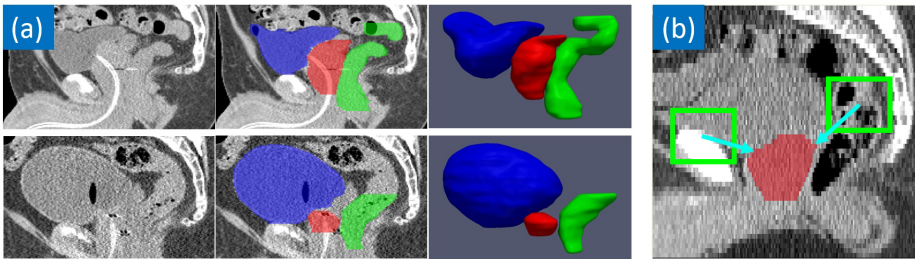

Fig. 1. (a) Two planning CTs and their segmentations of prostate (red), bladder (blue) and rectum (green). The 1st, 2nd, and 3rd columns show sagittal CT slices, the same slices overlaid with segmentations, and 3D views of segmentations, respectively. (b) 3D displacements (cyan arrows) to the organ boundary, where green boxes indicate the local patches of voxels.

organs from CT images. Chen et al. [2] used a Bayesian framework that integrates anatomical constraints from pelvic bones to segment the prostate and rectum. Costa et al. [3] proposed coupled deformable models to segment the prostate and bladder. Despite the popularity, deformable models are known to be sensitive to the initialization. If the initial shape model (e.g., 3D mesh) is not close to the target boundary, deformable models are likely to fail.

However, it is difficult to robustly initialize the shape model for CT pelvic organ segmentation due to two reasons: 1) pelvic organs show poor contrast (Fig. 1(a)), especially on their touching boundaries, and 2) the shape variations of pelvic organs are large across patients. These two factors hinder robust initialization, hence limiting the segmentation accuracy of deformable models.

In this paper, we propose a novel way to guide deformable models by jointly learning image regressor and classifier, which can greatly alleviate the problem caused by poor initialization. In the conventional learning-based deformable models, classifier is often used to guide deformable segmentation by producing an organ likelihood map, based on which each vertex of deformable model locally deforms along its normal direction to the position with the maximum boundary response. If the shape model is initialized far away from the true boundary, the conventional deformable models could suffer, as the local search will not be able to find the target boundary. To address this issue, in addition to classifier, we utilize random forest to jointly learn an image regressor. The image regressor predicts the 3D displacement from each image voxel to its nearest point on the organ boundary based on the local image appearance (Fig. 1(b)). It is used to drive each vertex of the deformable model towards the target boundary, thus relieving the problem caused by poor initialization. Once the deformable model is close to the boundary, the conventional local search strategy can be used to further refine the segmentation based on the organ likelihood map produced by the learned classifier. Validated on a CT dataset, experimental results show the robustness of our deformable model to the initialization as well as the effectiveness of our method in CT pelvic organ segmentation.

2 Method

2.1 Joint Learning of Image Regressor and Classifier

Random forest has been widely used in medical image analysis [1,5] due to its efficiency and effectiveness. In this paper, we use random forest to learn image regressor and classifier for each organ. The image regressor predicts the 3D displacement from any image voxel to its nearest point on the target organ boundary. The image classifier predicts the class label of each image voxel (i.e., target organ versus background). Both the image regressor and classifier make predictions according to the local appearance of image voxel. Considering that 3D displacement of a voxel is often correlated with its label, we propose to use random forest to learn image regressor and classifier jointly.

Joint Regression and Classification. Since both continuous variables (3D displacement) and discrete label need to be predicted together, we consider joint learning to be a joint regression and classification problem. Motivated by [5], we modify the objective function of random forest to consider both regression and classification in the tree optimization:

$$\{f^*, t^*\} = \max_{f,t}\{\frac{1}{Z_v}(v - \sum_{j\in\{L,R\}} \frac{N^j}{N} v^j) + \frac{1}{Z_c}(e - \sum_{j\in\{L,R\}} \frac{N^j}{N} e^j)\} \qquad (1)$$

where f and t are a feature and threshold, respectively. Here, 3D Haar-like features [4] are used to represent the local appearance of each voxel. v, e and N denote average variance, entropy and the number of training samples arriving at one node, respectively. The superscript $j \in \{L, R\}$ indicates the measurements computed after being split into the left or right child node. The first term of Eq. 1 computes variance reduction for regression, which is different from the differential entropy used in [5]. The second term of Eq. 1 computes entropy reduction for classification. Since the magnitudes of variance and entropy reductions are not of the same scale, we normalize both magnitudes by dividing the average variance and entropy at the root node (Z_v and Z_c), respectively.

Iterative Auto-Context Refinement. With the above modification, random forest can be learned to jointly predict the 3D displacement and class label for each voxel in the image. However, since both the displacement and label of each voxel are predicted independently from those of nearby voxels, the obtained 3D displacement field and label likelihood map are often spatially inconsistent, as shown in the first row of Fig. 2. To ensure the spatial consistency, it is necessary to consider the predictions of neighboring voxels during the voxel-wise displacement and label estimation.

Auto-context [6] is an iterative scheme that can incorporate the neighboring prediction information to compensate for the limitation of the independent voxel-wise estimation. The first iteration of training in the auto-context is the same as the conventional voxel-wise estimation method, which trains a prediction model according to local appearance features (e.g., Haar-like features) extracted from

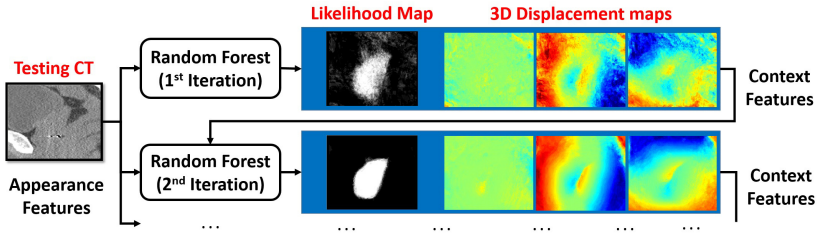

Fig. 2. Illustration of iterative auto-context refinement for prostate. Cold and warm colors in the color maps denote small and large distance values, respectively.

each voxel. In the second iteration, the model, learned in the previous iteration, is first applied to estimate the prediction maps (e.g., likelihood and displacement maps) for each training image. Additional features can be extracted at these prediction maps for each voxel. These features are named "*context features*" in the auto-context. Different from the traditional auto-context, which uses radiation-like context features, we extract Haar-like context features, which shows better performance. By combining appearance features from intensity image and context features from prediction maps, a new prediction model can be trained to update the prediction maps in the previous iteration. Following the same manner, a sequence of prediction models can be trained to iteratively update the prediction maps. Fig. 2 illustrates the testing stage of the auto-context.

Context features, which are extracted from local patches in previous prediction maps, capture the neighboring prediction information for a voxel, making this a significant area for improvement. As the predictions of neighboring voxels are highly correlated, the inclusion of context features in the prediction model could exploit this neighborhood correlation, which will improve the spatial consistency of prediction maps.

2.2 Deformable Segmentation with Regressor and Classifier

For each voxel in the image, the jointly learned image regressor and classifier can be used to estimate 3D displacement to the target organ boundary and organ likelihood. As introduced previously, the conventional learning-based deformable models use only the estimated organ likelihoods of voxels and rely on local search to drive deformable models onto the target organ boundary, causing them to be sensitive to model initialization. To overcome this problem, we utilize both image regressor and classifier to guide deformable models. Specifically, our deformable segmentation consists of two steps, as detailed below:

Pose Adjustment by Image Regressor. Considering that the deformable model may be initialized far from the target organ boundary, we first utilize image regressor to pull deformable model towards the boundary. Specifically, for each vertex of the deformable model, image regressor is used to estimate the 3D displacement of the vertex to the nearest point on the target organ boundary.

The estimated displacement is used as an external force to deform the shape model by sequentially adjusting its position, orientation and scale. Initially, the shape model is only allowed to translate under the guidance of image regressor. Once it is well positioned, we estimate its orientation by allowing it to rigidly deform. Finally, the deformation is relaxed to the affine transformation to estimate the scaling and shearing parameters of the shape model.

Boundary Refinement by Image Classifier. Once the shape model is close to the target organ boundary after pose adjustment, the conventional local search strategy can be used to refine the shape model based on the organ likelihood map provided by the image classifier. Specifically, in the refinement step, each vertex of deformable model locally searches along its normal direction to find its new position with the maximum likelihood gradient. After every vertex has deformed for one step, the entire shape model (3D mesh) is then smoothened prior to the next round of deformation. The vertex-wise deformation and mesh smoothing are iterated until convergence.

To increase the efficiency of our deformable model, we implement our model in multi-resolution. A joint image regressor and classifier is trained independently at two resolutions, such as coarse resolution and fine resolution. In the coarse resolution, the testing image is down-sampled by a factor of 4, which allows fast computation of rough segmentation. In the fine resolution, we need to compute only the 3D displacement field and organ likelihood map near the rough segmentation, which significantly reduces the computation time.

3 Experimental Results

Our dataset consists of 29 prostate cancer patients, each with one planning CT scan available. The prostate, bladder and rectum in all CT scans have been manually segmented by a physician, and serve as ground truth for our segmentation method. The image size is $512 \times 512 \times (400 \sim 500)$, and the image spacing is $1.172 \times 1.172 \times 1$ mm^3. Three-fold cross validation is used to evaluate our method. In random forest training, we prevent overfitting by stopping node splitting if the number of training samples arriving at one node is below 8. We do not limit the tree depth. The number of trees in the forest is 10. Three iterations are used in the auto-context.

Model Initialization. To demonstrate the robustness of our deformable model to the initialization, we initialize mean shape model at the center of every testing image. As shown in Fig. 3(a), the initialized shape model may be far away from the target organ. In such case, the conventional deformable models could fail, as the local search would not be able to find the interested organ boundary. In contrast, with the guidance from image regressor (Fig. 3(b)), the poorly initialized shape model could be effectively driven towards the target organ (Fig. 3(d)), which alleviates the dependency of deformable models on good initialization.

Auto-Context and Joint Model. Fig. 4 shows the magnitude of 3D displacement fields estimated by separately- and jointly- learned image regressors,

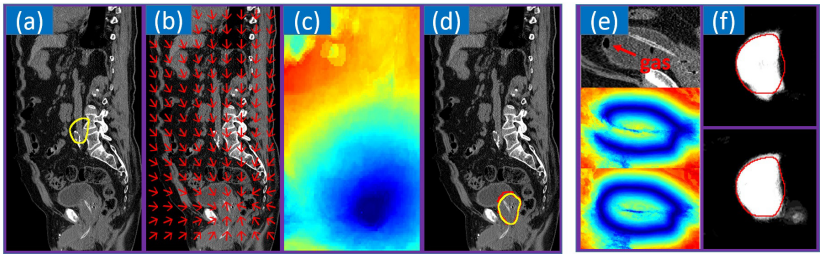

Fig. 3. (a-d) CT prostate segmentation. (a) Model initialization (yellow contour). (b-c) Direction and magnitude of 3D displacement field estimated in the coarse resolution. (d) Final segmentation. (e-f) Bladder displacement magnitude maps and prostate likelihood maps obtained by separate (upper) and joint (lower) models, respectively. Cold and warm colors in (c) and (e) denote small and large distances, respectively. Yellow contour in (d) denotes the automatic segmentation, and red contours in (d) and (f) denote ground truths.

without and with auto-context, respectively. We can see from Fig. 4(b-c) that the estimated 3D displacement field is noisy without auto-context, which makes it unreliable for guiding deformable models. We tested our method without auto-context, and found it failed in most cases. In contrast, with the help of auto-context, the quality of estimated 3D displacement significantly improves. Moreover, by comparing Fig. 4(d) and (e), we noticed that jointly-learned image regressor often obtains better displacement maps than separately-learned image regressor. In several cases (Fig. 3e), the displacement field estimated by separately-learned regressor may not necessarily form a closed surface (2nd row), due to inaccurate predictions for a small region. In such case, the pose adjustment will fail. In contrast, with evidence from classification, the jointly-learned regressor can effectively overcome this problem (3rd row), leading to better segmentation than separately-learned regressor (Table 1).

On the other hand, classification also benefits from regression. Fig. 3(f) gives one example of the prostate. Compared with separately-learned classifier (1st row), the jointly-learned classifier (2nd row) obtains better classification results on the boundary. This boundary difference also contributes to better segmentation accuracy of joint models compared to separated models (Table 1).

Table 2 (left) compares the traditional radiation-like context features with our Haar-like context features, which shows the superiority of Haar-like context features in the auto-context model.

Comparisons with Other Methods. Table 2 (right) compares our pose adjustment strategy with two other methods for model initialization. "Mass" uses the classification mass center from the coarse resolution, and "Box" uses the anatomy bounding box detection method [8]. The performances were obtained by replacing the pose adjustment strategy with respective initialization techniques. By comparing their results with ours in Table 1, we can conclude that the pose adjustment strategy is more effective in initializing deformable models of CT pelvic organs.

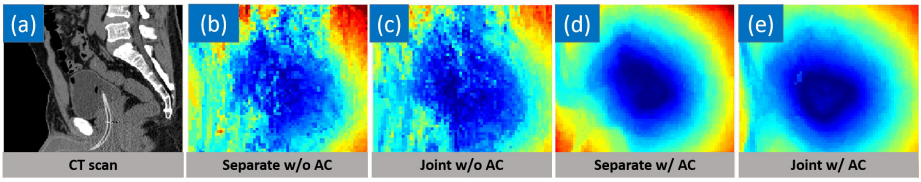

Fig. 4. Displacement magnitude maps of bladder obtained by different strategies. (b-c) Separately- and jointly- learned image regressors without auto-context. (d-e) Separately- and jointly- learned image regressors with auto-context.

Table 1. Quantitative comparison between separately-learned image regressor and classifier (Separate), and jointly-learned image regressor and classifier (Joint). Both of them use auto-context. DSC: Dice Similarity Coefficient for measuring overlap ratio. ASD: average surface distance for measuring boundary difference. Bold numbers indicate the best performance.

Organ		Prostate		Bladder		Rectum	
Resolution		Coarse	Fine	Coarse	Fine	Coarse	Fine
Separate	DSC(%)	72.7 ±11.3	84.8 ±3.97	79.0 ±15.3	92.1 ±13.3	70.9 ±7.05	84.3 ±6.25
	ASD(mm)	4.19 ±1.60	2.28 ±0.55	4.15 ±2.12	1.38 ±1.02	3.23 ±0.69	1.97 ±0.97
Joint	DSC(%)	73.0 ±8.99	**85.2 ±3.74**	82.7 ±5.74	**94.9 ±1.62**	69.3 ±7.52	**84.7 ±5.22**
	ASD(mm)	4.20 ±1.32	**2.23 ±0.53**	3.78 ±1.26	**1.15 ±0.29**	3.51 ±0.82	**1.97 ±0.92**

Table 3 quantitatively compares our method with other existing segmentation methods. Since different methods use different metrics to measure their performance, we separate the comparisons into two tables. We can see from Table 3 that our method achieves better accuracy than the existing methods under comparison. It is worth noting that most previous works use either sophisticated methods for model initialization [1,3,7], or rely on shape priors to regularize the segmentation [1,2,3,7]. In contrast, our method uses a fairly simple initialization method (i.e., initialize the mean shape model at image center), and does not rely on shape priors (e.g., PCA shape analysis, or sparse shape model [9]). It is interesting to observe that even with this setup, our method still results in more competitive results, when compared to previous methods. This demonstrates the robustness of our method in the initialization and effectiveness of our method for CT pelvic organ segmentation.

Table 2. Quantitative comparisons (left) between radiation-like (RADI) and Haar-like (Haar) context features in the auto-context, and (right) between different deformable initialization methods, where "Mass" uses the classification mass center from the coarse resolution to initialize deformable models.

DSC(%)	Prostate	Bladder	Rectum
RADI	84.5±4.4	94.0±6.1	82.7±8.5
Haar	85.2±3.7	94.9±1.6	84.3±6.2

DSC (%)	Prostate	Bladder	Rectum
Mass	81.5±12.0	82.9±18.1	46.0±26.8
Box[8]	74.4±21.2	71.7±26.0	66.4±20.3

Table 3. Comparisons of our method with other existing works. Left panel shows the comparison based on average surface distance (ASD). Right panel shows the comparison based on sensitivity (Sen) and positive predictive value (PPV). Good segmentation results have small ASD, and large Sen and PPV.

	Prostate	Bladder	Rectum
Mean±Std	ASD(mm)	ASD(mm)	ASD(mm)
Lay [1]	3.57±2.01	2.59±1.70	4.36±1.70
Lu [7]	2.37±0.89	2.81±1.86	4.23±1.46
Proposed	**2.23±0.53**	**1.15±0.29**	**1.97±0.92**

	Prostate		Bladder		Rectum	
Median	Sen	PPV	Sen	PPV	Sen	PPV
Costa [3]	0.79	0.86	0.87	0.93	N/A	N/A
Chen [2]	**0.84**	0.87	N/A	N/A	0.71	0.76
Proposed	0.82	**0.91**	**0.94**	**0.97**	**0.88**	**0.84**

4 Conclusion

In this paper, we propose a novel way to guide deformable segmentation by jointly learning image regressor and classifier for each organ. The image regressor is used to initially drive deformable model towards the organ boundary, while the image classifier is used to refine the segmentation once the pose (i.e., position, orientation and scale) of the deformable model has been adjusted by image regressor. Experimental results, based on a CT pelvic dataset, demonstrate the robustness of our deformable model to initialization, as well as the competitive performance of our method to other existing methods.

References

1. Lay, N., Birkbeck, N., Zhang, J., Zhou, S.K.: Rapid multi-organ segmentation using context integration and discriminative models. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) IPMI 2013. LNCS, vol. 7917, pp. 450–462. Springer, Heidelberg (2013)
2. Chen, S., Lovelock, D.M., Radke, R.J.: Segmenting the prostate and rectum in CT imagery using anatomical constraints. Medical Image Analysis 15(1), 1–11 (2011)
3. Costa, M.J., Delingette, H., Novellas, S., Ayache, N.: Automatic segmentation of bladder and prostate using coupled 3D deformable models. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 252–260. Springer, Heidelberg (2007)
4. Gao, Y., Wang, L., Shao, Y., Shen, D.: Learning distance transform for boundary detection and deformable segmentation in CT prostate images. In: Wu, G., Zhang, D., Zhou, L. (eds.) MLMI 2014. LNCS, vol. 8679, pp. 93–100. Springer, Heidelberg (2014)
5. Glocker, B., Pauly, O., Konukoglu, E., Criminisi, A.: Joint classification-regression forests for spatially structured multi-object segmentation. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part IV. LNCS, vol. 7575, pp. 870–881. Springer, Heidelberg (2012)
6. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. TPAMI 32(10), 1744–1757 (2010)

7. Lu, C., Zheng, Y., Birkbeck, N., Zhang, J., Kohlberger, T., Tietjen, C., Boettger, T., Duncan, J.S., Zhou, S.K.: Precise segmentation of multiple organs in CT volumes using learning-based approach and information theory. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 462–469. Springer, Heidelberg (2012)
8. Criminisi, A., et al.: Regression forests for efficient anatomy detection and localization in computed tomography Scans. MedIA 17(8), 1293–1303 (2013)
9. Zhang, S., et al.: Towards robust and effective shape modeling: Sparse shape composition. MedIA 16(1), 265–277 (2012)