Semi-Automatic Prostate Segmentation Using Semi-supervised And Active Learning

Dwarikanath Mahapatra, Joachim M. Buhmann

Department of Computer Science, ETH Zurich, Switzerland.

Abstract. We propose an interactive active learning (AL) approach for prostate segmentation from magnetic resonance (MR) images. A VOI is manually defined and its probability map is generated using a Random forest (RF) classifier. The RF is trained using semisupervised learning (SSL) and informative labels are queried using active learning and random walks. Graph cuts are used to obtain the final segmentation. Our method was applied to the PROMISE12 Prostate segmentation dataset. The classifier was trained on the training dataset consisting of 50 patient volumes and then applied to the test data comprising 30 volumes.

1 Introduction

Prostate cancer is one of the leading cause of male cancer death in the USA [1]. Automatic prostate segmentation from magnetic resonance (MR) images is important for radiotherapy, prostate volumetry and calculation of prostate specific antigen (PSA) density. However it faces challenges due to: 1) variability of prostate size and shape between subjects; 2) different MR scanning protocols resulting in different image appearances and intensity ranges; 3) the lack of clear prostate boundaries due to similar intensity profiles of surrounding tissues.

Prostate segmentation from computed tomography (CT) and Ultrasound (US) images have been widely investigated [2,3]. MRI has gained popularity because of lack of ionising radiations as in the case of CT images and better contrast between soft tissue organs than US scans. However, accurate delineation of the prostate's border remains difficult even for the human observer.

Klein et al. [4] proposed a multi-atlas matching for prostate segmentation using localized mutual information. Toth et al. [5] used shape prior deformable models while Li et al. [6] used auto context and level-sets to achieve binary segmentation of the prostate. In the MICCAI 2012 prostate segmentation challenge, different approaches used learning techniques [7], multi-atlases [8], active shape models (ASM) [9] and active appearance models (AAM) [10,11] to name a few.

Recent research has focused on learning highly discriminative image features from pre-annotated images. A robust algorithm requires many example cases to learn from a wide range of image appearances. However, obtaining manual annotations is very expensive, time consuming and requires personnel with high expertise. We propose an interactive machine learning (ML) based segmentation method that requires significantly fewer labeled samples, yet achieves higher

segmentation accuracy than conventional ML methods. Our algorithm uses 1) an active learning (AL) strategy that queries the labels of informative regions; and 2) semi-supervised learning (SSL) that uses a few labeled samples and many unlabeled samples to construct a highly accurate classifier.

2 Image Features

The features are extracted from a 31×31 patch around each voxel, while the query patches are 8×8. Intensity, texture and curvature features combined give a 85 dimensional feature vector, while context features give an additional 96 features. Since our emphasis in this paper is the combination of SSL and AL techniques, we give a brief description of each feature and refer the reader to [12] for details.

2.1 Intensity Statistics

MR images commonly contain regions that do not form distinct spatial patterns but differ in their higher order statistics. Therefore, in addition to the features processed by the human visual system (HVS), i.e., mean and variance, we extract skewness and kurtosis values from each voxel's 31×31 neighborhood.

2.2 Texture Entropy

Texture maps are obtained from 2-D (instead of 3D) Gabor filter banks for each slice (at orientations $0°, 45°, 90°, 135°$ and scale $0.5, 1$). Each 31×31 image patch is partitioned into 9 equal parts corresponding to 9 sectors of a circle. Figure 1 (a) shows the template for partitioning a patch into sectors and extracting entropy features. For each sector we calculate the texture entropy given by

$$\chi^r_{ani} = -\sum_{tex} p^r_{tex} \log p^r_{tex}. \tag{1}$$

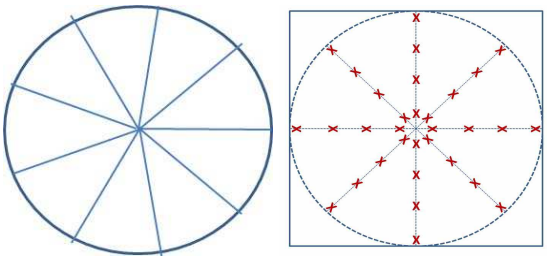
p^r_{tex} denotes the probability distribution of texture values in sector r. This procedure is repeated for all the 8 texture maps over 4 orientations and 2 scales to extract a $(8 \times 9 =)$ 72 dimensional feature vector.

2.3 Curvature Entropy

Different tissues have different curvature distributions and we exploit this characteristic for accurate prostate identification. Details on curvature calculation is given in [12]. Similar to texture, curvature entropy is calculated from 9 sectors of a patch and is given by

$$Curv^r_{ani} = -\sum_{\theta} p^r_{\theta} \log p^r_{\theta}. \tag{2}$$

p^r_{θ} denotes the probability distribution of curvature values in sector r, and θ denotes the curvature values.



(a) (b)

Fig. 1. (a) partitioning of patch for calculating anisotropy features; (b) template for calculating context features.

2.4 Spatial Context Features:

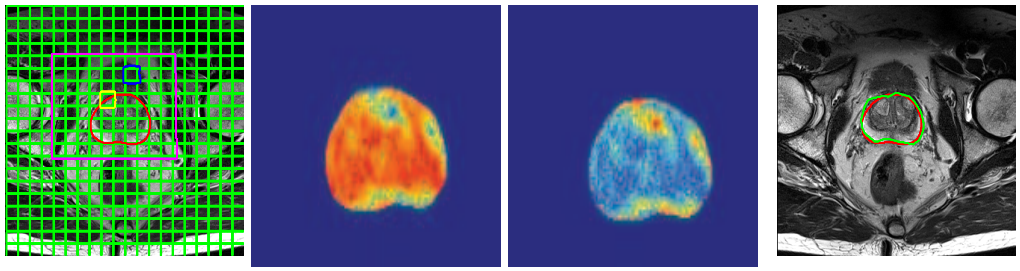
Context information is particularly important for medical images because of the the regular arrangement of human organs. Figure 1 (b) shows the template for context information where the circle center is the current voxel and the sampled points are identified by a red 'X'. At each voxel corresponding to a 'X' we calculate the mean intensity, texture and curvature values from a 3×3 patch. The texture values were derived from the texture maps at 90° orientation and scale 1. The 'X's are located at distances of $1, 3, 5, 8$ voxels from the center, and the angle between consecutive rays is 45°. The values from the 32 regions are concatenated into a 96 dimensional feature vector, and the final feature vector has $(96 + 85 =)181$ values.

3 Methods

3.1 Initial Preprocessing

The given images are first bias-corrected using the method in [13] to remove intensity inhomogeneities due to the magnetic field of MR machines. The intensities are then normalized to $[0, 1]$. Pixels with the lowest 5% intensities usually indicate noise, and are all set to zero. Similarly, higher intensity tails in the intensity histogram indicate artifacts and outliers. Hence the mean of the 5% brightest intensity values is taken to be the maximum image intensity. All intensity values are divided by this threshold and ratios exceeding one are set to one. Normalization also increases the image contrast. A manual VOI is drawn by the user and all subsequent computations are restricted to this VOI. This helps: 1) save time; and 2) to reduce false positives in classification since the background has regions with widely varying intensity profiles.

A given test image is divided into non-overlapping uniform 8×8 patches and an expert identifies one patch each belonging to prostate and background. The voxels of each patch now have the same labels. Features (described in Section 2) are extracted for each *voxel*. The features of labeled and unlabeled patches are



(a) (b) (c) (d)

Fig. 2. (a) Original image divided into patches (green). Colored patches show prostate (yellow), background (blue) and VOI (magenta). Manual ground truth segmentation is in red; Probability maps for (b) prostate and (c) background. Red indicates maximum value of 1 and blue indicates 0; (d) final segmentation obtained by our method in green and manual segmentation in red.

used as inputs to an RF based SSL classifier (denoted as $RF - SSL$). $RF - SSL$ outputs the class labels and probabilities of the unlabeled voxels. We are interested only in the class probabilities for determining the most informative patch for label query.Note that the designated probability value for a patch is the average of its constituent pixels.

The label information of the queried sample is used to update the classifier using the online learning strategy of [14] and predict a new set of labels. Figure 2 (a) shows the original image and the patches as shown to the expert. The patches selected by the clinician as prostate (yellow) and background (blue), and the VOI of the slice (magenta) are also shown. Manually annotated prostate is shown in red. Note that the manually annotated regions are not shown to the clinician and are used only for evaluating the accuracy of our algorithm. Figures 2 (b,c) show the VOI probability maps for prostate and background, followed by our segmentation in Figure 2 (d) (green).

3.2 AL Query Strategy

Our query strategy selects a sample with the following characteristics: 1) high classification uncertainty to obtain novel information from each labeling instance; 2) situated in a dense region such that it is representative of many other samples; and 3) minimal overlap of influence from previously labeled samples to minimize redundancy in labeling effort.

Choosing the query sample is similar selecting the most salient region of an image based on some measure of informativeness. Salient image regions have the following characteristics: 1) their feature values are significantly different from surroundings (high local contrast); and 2) contrast magnitude is higher than other regions and hence it stands out visually.

High contrast regions have maximum information and hence higher entropy [15]. Regions with high classification uncertainty also have high entropy, indi-

cating a correspondence between information content of salient regions and classification uncertainty. Salient regions are located on regular objects (or dense regions of the sample space) and different salient regions are far away from each other, i.e. their influence areas have minimum overlap. Thus we see that the properties of salient regions have similarities with the desired characteristics of query samples. Hence saliency models can be adapted for active learning tasks using appropriate similarity metrics.

Image patches are represented as nodes V of a graph G, and connected by set of edges E. Based on the similarity between any two nodes i and j a weight w_{ij} is assigned to edge e_{ij}. Random walks on the graph determine the most frequently visited node which is also the most salient node. For further details we refer the reader to [16,17]. Informativeness of node i (or patch x) is given by

$$Inf(i) = \{\phi(i),\ A, \alpha\}\,. \tag{3}$$

ϕ is the classification uncertainty of i given by the entropy as

$$\phi(i) = -\sum_{\widehat{y}} P((\widehat{y}|i) \log P((\widehat{y}|i)\,, \tag{4}$$

where $\widehat{y}$ indicates all possible labels (in this case two) for i, and $P((\widehat{y}|i)$ is calculated by RF-SSL. High entropy indicates greater uncertainty. α incorporates contextual information, and A is the collection of intensity, texture and curvature differences defined as

$$A = [Int_{ij}\ Tex_{ij}\ Curv_{ij}]. \tag{5}$$

where $Int_{ij} = \sum_{j \in N} e^{-|Int_i - Int_j|/\sigma^2}$ is the sum of exponential of intensity differences between node i and all unlabeled nodes j in N (a 48×48 neighborhood of x), and $\sigma = 1$. For similar nodes, Int_{ij} takes higher values. Tex_{ij} and $Curv_{ij}$ are the corresponding texture (from the oriented map at 90°) and curvature differences. Note that we do not average the feature differences over the neighborhood. In a high density region A is calculated by summing over more voxels than in a sparsely populated region. Since A is not divided by a normalization constant its value is higher in a high density region.

Context Information for Informativeness: Medical images have inherent context information because the relative arrangement of organs in the human body is constant, and similar tissues are clustered together. An unlabeled sample close to a labeled sample is assigned lower importance because it has a higher probability of having the same label than a sample far away. With every annotation we want to derive maximum information about its neighbors. If the radiologist were to annotate samples close to an already labeled sample it *does not* generally lead to significant information gain. Thus α incorporates context information and is equal to i's distance from the nearest labeled sample

$$\alpha = \min\left(\left\|i - i^L\right\|\right). \tag{6}$$

where i^L denotes all the labeled samples (or nodes), and $\|.\|$ denotes the Euclidean distance based on voxel co-ordinates.

Thus the feature vector of node i (F_i or the informativeness $Inf(i)$) consists of 5 values i.e., entropy, intensity, texture and curvature differences, and α. The weights between two nodes i and j are given by

$$w_{ij} = \exp\left(\frac{-\|F_i - F_j\|^2}{\sigma^2}\right), \tag{7}$$

where $\sigma = 1$, and $\|.\|$ denotes L_2 norm. Nodes with similar features have higher connecting weights.

Thus we observe that Eqn. 4 encodes the classification uncertainty of a node (or image patch), Eqn. 5 quantifies the patch's density and Eqn. 6 determines whether a patch is situated within the region of influence of an already labeled sample. Note that for ϕ and A the value of the patch is the average of its constituent voxels, and α is determined by the distance between the center voxels of each patch. The corresponding elements of F_i and F_j have values in the same range. If the two nodes i and j are similar the differences of each element is also in the same range. The query node has high differences with its neighboring nodes and is thus the most salient node as discussed previously.

3.3 Random Walks and Most Salient node

Current saliency detection algorithms incorporate more sophisticated conditions for salient region detection that takes into consideration global image characteristics. Similarly active learning based query detection can be significantly improved by using saliency algorithms that incorporate more global image information (such as graph based methods).

The most salient node is identified by the random walks algorithm on the graph. Let us denote as $E_i(T_j)$ the expected number of steps to reach state j if a Markov chain is started in state i at time $t = 0$. It is also known as the hitting time, and can be derived from the fundamental matrix ($\mathbf{Z}$) of an ergodic Markov chain and its equilibrium probability distribution π. The global saliency of node i is given by the sum of hitting times from all other nodes to node i on a complete graph, i.e.,

$$H_i = \sum_j E_j(T_i), \tag{8}$$

and the most salient node is given by the maximum H_i as $N_s = \arg\max_i H_i$. For details the reader is referred to [17]. Labels are queried for N_s.

Stopping Criteria: After every classification we determine the distribution of probability values for a single class and compare with the corresponding values in the previous iteration using the Student $t-$test. $p < 0.05$ denotes statistically different classification in the current iteration. However, $p > 0.05$ indicates minor difference in the distributions. If $p > 0.05$ for two consecutive iterations there is no further querying of labels.

3.4 Graph Cut Segmentation

A spatially smooth solution is obtained by formulating the segmentation as a labeling problem within a second order MRF cost function. The labels are obtained by optimizing the cost function using graph cuts [18]. The labels are obtained for each voxel and not for individual patches. A second order MRF energy function is given by

$$E(L) = \sum_{s \in P} D(L_s) + \lambda \sum_{(s,t) \in N} V(L_s, L_t), \tag{9}$$

where P denotes the set of pixels and N is the set of neighboring pixels for pixel s. λ is a weight that determines the relative contribution of penalty cost (D) and smoothness cost (V). $D(L_s)$ is given by

$$D(L_s) = -\log\left(Pr(L_s) + \epsilon\right), \tag{10}$$

where Pr is the likelihood (from probability maps) previously obtained using RF classifiers and $\epsilon = 0.00001$ is a very small value to ensure that the cost is a real number. The penalty cost encourages high label probability.

V ensures a spatially smooth solution by penalizing discontinuities. Let the normalized weight (importance measures) of the different features be w_I (intensity), w_T (texture) and w_C (curvature), where $w_I + w_T + w_C = 1$. V is given by

$$V(L_s, L_t) = \begin{cases} w_I V_I + w_T V_T + w_C V_C, & L_s \neq L_t, \\ 0 & L_s = L_t. \end{cases} \tag{11}$$

where V_I, V_T, V_C are the individual contributions to the smoothness by intensity, texture and curvature. V_I is defined as

$$V_I(L_s, L_t) = e^{-\frac{(I_s - I_t)^2}{2\sigma^2}} \cdot \frac{1}{\|s - t\|}, \tag{12}$$

I denotes the intensity. V_T and V_C are similarly defined using texture and curvature instead of intensity. Smoothness cost is determined over a 8 neighborhood system.

4 Experiments and Results

As part of the MICCAI 2012 prostate segmentation challenge (http://promise12.grand-challenge.org/) the training dataset consists of 50 transversal T2-weighted MR datasets of the prostate. The dataset is a representative set of the types of MR images acquired in a clinical setting. The data is multi-center and multi-vendor and has different acquisition protocols (e.g. differences in slice thickness, with/without endorectal coil). The set is selected such that there is a spread in prostate sizes and appearance. Reference segmentations are available for each dataset. We employ a 10 fold cross validation where training is done on 45

datasets and tested on 5 datasets. We report average test results over all patients. N3 intensity non-uniformity correction was applied to reduce intensity inhomogeneity and the image intensities are normalized using the method in [19]. Table 2 gives details about different aspects of our algorithm. The quality of our segmentations was evaluated using two measures: 1) Dice Metric (DM) and 2) Hausdorff Distance (HD).

4.1 Segmentation Results

We present comparative results on the training data for:

1. $SSL - AL$: Our proposed segmentation algorithm using RF, SSL and AL.;
2. FSL: our automatic fully supervised learning based prostate segmentation of [20] that does not use active learning. A 5 fold cross validation approach is used.
3. $SSL - AL_{nV}$: $SSL - AL$ without incorporating variable importance measures in smoothness cost V.
4. $SSL - AL_{CV}$: $SSL - AL$ with 5 fold cross validation where the classifier is trained using SSL and AL on the training data and the trained classifier used to segment the test images without expert feedback. For segmenting the the test data we use the manually defined VOI for $SSL - AL$ and then segment the prostate. Our segmentation results on the test set are obtained by applying $SSL - AL_{CV}$.

Fig. 3 shows segmentation results on Patient 24 using the above methods. Table 1 summarizes the average segmentation results over all patients for different methods. $SSL - AL$ gives the best segmentation results, and significantly improves over FSL's performance with $p < 0.001$. $SSL - AL$ achieves higher segmentation accuracy using fewer labeled samples, which translates to less labeling time. $SSL - AL_{nV}$ gives lower segmentation accuracies than $SSL - AL$ because of the absence of importance measures in V. This is particularly important while segmenting regions of low contrast whose appearance is similar to its neighbors.

$SSL - AL_{CV}$'s performance is very close to that of $SSL - AL$ with $p > 0.1$. This indicates that SSL and AL combine nicely to give a robust classifier even if it does not have access to the labels of the target scans in terms of expert feedback. Importantly, $SSL - AL_{CV}$ is able to select the right samples to learn from and build a robust classifier.

$SSL - AL_{CV}$ takes significantly less time because of the manually defined VOI due to which it does not have to automatically locate the VOI. Since there is no expert feedback for $SSL - AL_{CV}$ computation time is further reduced. Although $SSL - AL_{CV}$'s computation time is less its performance is very similar to $SSL - AL$.

5 Conclusion

We have presented a interactive method for prostate segmentation using active learning and semi supervised learning. Semi supervised learning makes optimal

Table 1. Quantitative measures for segmentation accuracy. DM-Dice Metric in %, HD-Hausdorff distance in mm, Time-computation time in minutes.

	$SSL-AL$	FSL	$SSL-AL_{nV}$	$SSL-AL_{CV}$
DM	91.9	90.2	87.3	91.3
HD	6.2	7.9	10.3	6.8
Time(min)	12.3	21.2	12.6	10.9

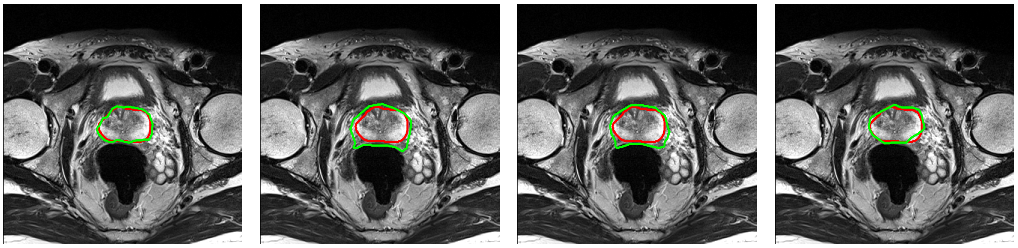


Fig. 3. Segmentation results for Training Patient 24. The manual annotations are shown in red with the algorithm segmentations in green: $SSL-AL$ (Column 1), FSL (Column 2), $SSL-AL_{nV}$ (Column 3), and $SSL-AL_{CV}$ (Column 4). Each row shows results for different slices of the same volume.

use of labeled and unlabeled data while active learning makes optimal use of annotations by lowering their redundancy. Active learning queries labels of samples that provide maximal information gain, and the query strategy is based on: 1) high classification uncertainty by the SSL classifier; 2) its location in a dense region such that it is representative of many other samples; and 3) minimizing overlap of region of influence of every labeled sample such that optimal classification is achieved using minimum number of samples. We obtain higher segmentation accuracy than FSL methods with less than 50% samples, a significant reduction in labeling effort.

References

1. : Cancer society atlanta:www.cancer.org,2011
2. Chen, S., Lovelock, D., radke, R.: Segmenting the prostate and rectum in ct imagery using anatomical constraints. Med. Imag. Anal. **15**(1) (2011) 1–11
3. Zhan, Y., Shen, D.: Deformable segmentation of 3-d ultrasound prostate images using statistical texture matching method. IEEE Trans. Med Imag. **25**(3) (2006) 256–272
4. Klein, S., van der Heide, U., Lipps, I., Vulpen, M., Staring, M., Pluim, J.: Automatic segmentation of the prostate in 3d mr images by atlas matching using localized mutual information. Medical Physics **35**(4) (2008) 1407–1417
5. Toth, R., Tiwari, P., Rosen, M., Reed, G., Kurhanewicz, J., Kalyanpur, A., Pungavkar, S., Madabhushi, A.: A magnetic resonance spectroscopy driven initializa-

Parameter	Value
Language	MATLAB
Libraries/Packages	Image Processing toolbox, Graph cut optimization, Random Forest Classifiers
Multi Threaded	No
User Interaction	Yes (for selecting training samples)
CPU Clock Speed	2.66 GHz
Machine CPU Count	2
Machine Memory	4 GB
Memory Used during Segmentations	< 1 GB
Training Time (Voxel Classifier)	45 minutes
Image Preprocesing time	1-2 min
Voxel Classification + Segmentation Time	7-9 mins
Total Segmentation Time	9-11 mins

Table 2. Parameter values for segmentation as required in $SSL - AL$

tion scheme for active shape model based prostate segmentation. Medical Image Analysis **15**(2) (2011) 214–225

6. Li, W., Liao, S., Feng, Q., Chen, W., Shen, D.: Learning image context for segmentation of prostate in ct-guided radiotherapy. In: MICCAI. (2011) 570–578
7. Ghose, S., Mitra, J., Oliver, A., Marti, R., Llado, X., Freixenet, J., Vilanova, J., Sidibe, D., Meriaudeau, F.: A random forest based classification approach to prostate segmentation in mri. In: Proc. MICCAI Prostate Segmentation Challenge. (2012)
8. Litjens, G., Karssemeijer, N., Huisman, H.: A multi-atlas approach for prostate segmentation in mr images. In: Proc. MICCAI Prostate Segmentation Challenge. (2012)
9. Kirschner, M., Jung, F., Wesarg, S.: Automatic prostate segmentation in mr images with a probabilistic active shape model. In: Proc. MICCAI Prostate Segmentation Challenge. (2012)
10. Maan, B., van der Heijden, F.: Prostate mr image segmentation using 3d active appearance models. In: Proc. MICCAI Prostate Segmentation Challenge. (2012)
11. Vincent, G., Guillard, G., Bowes, M.: Fully automatic segmentation of the prostate using active appearance models. In: Proc. MICCAI Prostate Segmentation Challenge. (2012)
12. Mahapatra, D., Schüffler, P., J.Tielbeek, Makanyanga, J., Stoker, J., Taylor, S., Vos, F., Buhmann, J.: Automatic detection and segmentation of crohn's disease tissues from abdominal mri. IEEE Trans. Med. Imaging **32**(12) (2013) 2332–2348
13. Cohen, M., Dubois, R., Zeineh, M.: Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. Human Brain Map. **10**(4) (2000) 204–211
14. Saffari, A., Leistner, C., Santner, J., Godec, M., Bischof., H.: On-line random forests. In: IEEE ICCV Workshops. (2009) 1393 – 1400
15. Kadir, T., Brady, M.: Saliency, scale and image description. International journal of Computer Vision **45**(2) (2001) 85–105
16. Gopalakrishnan, V., Hu, Y., Rajan, D.: Random walks on graphs to model saliency in images. In: IEEE CVPR. (2009) 1698–1705

17. Harel, J., Koch, C., Perona, P.: Graph based visual saliency. In: NIPS. (2006) 545–552
18. Boykov, Y., Veksler, O.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23** (2001) 1222–1239
19. Nyl, L., Udupa, J.: On standardizing the mr image intensity scale. Magnetic resonance in medicine **42**(6) (1999) 1072–1081
20. Mahapatra, D., Buhmann, J.: Prostate mri segmentation using learned semantic knowledge and graph cuts. IEEE Trans. Biomed. Engg. **61**(3) (2014) 756–764