Adrenal Lesions Detection on Low-Contrast CT Images using Fully Convolutional Networks with Multi-Scale Integration

Lei Bi[1], *Jinman Kim*[1], *Tingwei Su*[2], *Michael Fulham*[1,3,4], *Dagan Feng*[1,5], and *Guang Ning*[2]
[1]School of Information Technologies, University of Sydney, Australia
[2]Ruijin Hospital, Shanghai Jiao Tong University, China
[3]Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Australia
[4]Sydney Medical School, University of Sydney, Australia
[5]Med-X Research Institute, Shanghai Jiao Tong University, China

ABSTRACT

Adrenal lesions include a wide variety of benign and malignant neoplasms of the adrenal gland, and are seen in up to 5% of computed tomography (CT) examinations of the abdomen. Better identification of these lesions is important for effective management and patient prognosis. Detection on low-contrast CT images, however, even for experienced physicians can be difficult and error-prone, because the lesions are often problematic to be separated from the normal surrounding structures. Existing lesion detection techniques have problems in identifying and differentiating low-contrast tumors, which is related to their use of low-level features rather than high level of semantics. Hence we propose an automated approach using fully convolutional networks (FCNs) and multi-scale integration to detect adrenal lesion on low-contrast CT scans. The architecture of FCNs includes deep, coarse, semantic information and shallow, fine, appearance information in a hierarchical manner and it enables the encoding of image-wide location and semantics, which are desirable characteristics for adrenal lesion detection. We also propose a multi-scale integration with a superpixel based random walk (MI-SRW) approach to refine the lesion boundaries on different scales. The MI-SRW technique enables us to constrain the spatial and appearance consistency and then use complementary information provided on different scales to detect adrenal lesions of various sizes and characteristics. We used 38 adrenal lesions detected on low-contrast CT and compared our approach to existing 'state-of-the-art' methods and found that our approach had superior detection performance.

Index Terms— Detection, Fully Convolutional Networks (FCN), Adrenal lesions

1. INTRODUCTION

Adrenal lesions consist of benign and malignant neoplasms that affect the adrenal gland, and include functional adenomas, pheochromocytomas, adrenal carcinomas and metastases [1]. These lesions are commonly seen on abdominal CT studies (approx. 5%) and also seen as incidental findings on CTs [2-5]. Better identification of these lesions is important for effective management and patient prognosis. The detection, however, of these lesions on non-contrast enhanced CT is problematic even for experienced physicians. This relates to variations in size, shape and lesion intensity. In addition, these lesions usually exhibit similar characteristics to normal adjacent structures including liver, bowel and kidneys (see Figure 1). Computer-aided diagnosis (CAD) systems are a useful '2nd opinion' tool for physicians in image interpretation, however, to our knowledge, CAD has not been used in detecting adrenal lesions on low contrast CT.

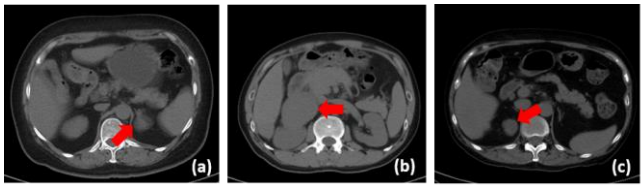

Figure 1. Examples (a-c) of adrenal lesions (red arrow) on CT images.

Existing image processing techniques for tumor detection mainly focus on detecting tumors where there is a high lesion-to-background contrast e.g., non-small cell lung cancers [5] or in identifying lesions within a solid organ such as the liver [6]. These techniques also generally use low-level features e.g., texture features extracted from a small patch (superpixel) and a classifier such as a support vector machine (SVM). Hence they are not suitable for adrenal lesions where the lesion may have similar intensity to adjacent normal structures. Recently, fully convolutional networks (FCNs) have achieved a great success in object detection related problems via their use of high-level semantics [7, 8]. Further, unlike existing methods that often depend on many hand-engineered pre-processing steps, FCNs can be trained with efficient inference of learning in an end-to-end fashion, which takes an image as input and directly outputs a probability map. The FCN architecture combines deep, coarse, semantic information and shallow, fine, appearance information in a hierarchical manner that enables it to encode

image-wide location and semantics, which are desirable properties for lesion detection on low-contrast CT images. However, FCNs have large receptive fields in the convolutional filters and hence produce coarse outputs when resized to probability maps. FCNs also lack smoothness constraints to encourage label agreement among neighboring similar pixels and therefore, it is difficult to produce a probability map with spatial and appearance consistency. Hence, we propose a fully automated FCN-based approach with multi-scale integration to address these limitations. The novelty of our approach, when compared to previous research, is as follows: (1) our FCN-based approach takes CT image as input and outputs the lesion probability map directly and does not require any hand-crafted or pre-processing steps; (2) we propose a multi-scale integration with superpixel based random walk (MI-SRW) technique to enhance the adrenal lesion boundaries and to integrate the detected results on different scales. The MI-SRW technique can deal with lesions of varying sizes, encourages label agreement among neighboring pixels and constrains the spatial and appearance consistency.

2. METHODS

2.1. Fully Convolutional Networks (FCNs) for Lesion Detection

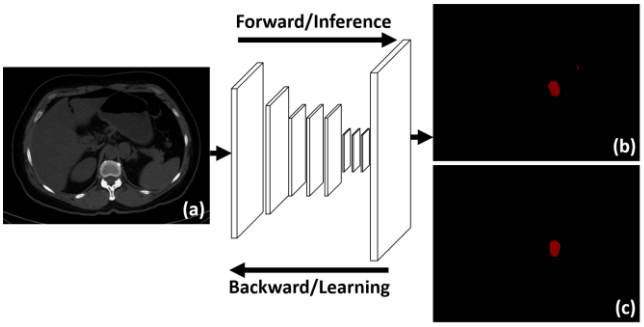

Figure 2. A FCN for lesion detection: (a) input transaxial CT; (b) FCN results; and (c) the ground truth of input CT image (image (a)).

The FCN architecture contains downsampling and upsampling parts [7]. The downsampling part has convolutional and max-pooling layers to extract high level abstract information and has been widely used in convolutional neural networks (CNN) for image classification task [10]. The upsampling part has convolutional and deconvolutional layers that upsample the feature maps and output the score masks. The FCN, by combining downsampling and upsampling parts, can extract the high-level semantic information while predicting the pixel-wise score mask. Currently, there is a relative scarcity of medical image training data with annotations due to cost and complicated acquisition procedures [9]. Consequently, without sufficient training data to cover all the adrenal lesions with various contrast and size variations, FCN cannot provide accurate detection and tumor boundary definitions. There are, however, extensive research shows that fine-tuning the deep convolutional neural networks (DCNNs) can alleviate the problem of insufficient medical training data [9]. Therefore, we used the off-the-shelf MatConvNet [11] version of FCN trained on the PASCAL VOC 2011 dataset. We used data augmentation techniques including random crops and contrast variations to provide extra material for training. We fine-tuned a stride-8 FCN architecture (FCN-8s) on the adrenal lesion images for more precise details of pixel predication. We fine-tuned the network for 150 epochs with a batch size of 20. The FCN converged at approximately 100 epochs.

2.2. Multi-Scale Integration with Superpixel based Random Walk (MI-SRW)

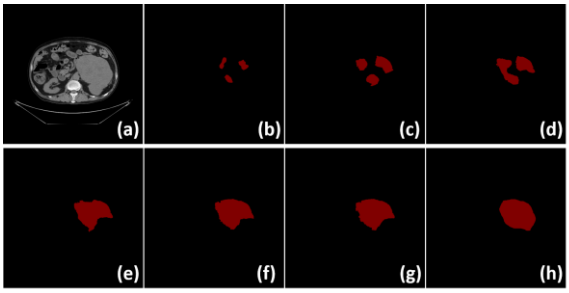

Figure 3. Superpixel based random walk: (a) input transaxial CT; (b) FCN results; (c-g) refined results of (b) at different scales with random walk (from 200 to 1000 superpixels with an increment of 200 per image); and (h) ground truth.

From the FCN, we can roughly estimate the location of the adrenal lesions. These detected lesions, however, are usually coarse and display large semantic differences to the actual ground truth (Figure 3b); this is due to the large receptive fields and lack of constraint on neighboring pixel agreement. So we introduce the multi-scale integration with superpixel based random walk (MI-SRW) to refine the lesion delineation. Given a list of superpixels, the idea of superpixel based random walk is to minimize the Dirichlet integral [12], defined as:

$$Dir[p^k] = \frac{1}{2}(p^k)^{\mathrm{T}} L(p^k) = \frac{1}{2}[(p_s^k)^{\mathrm{T}} \quad (p_u^k)^{\mathrm{T}}] \begin{bmatrix} L_s & B \\ B^{\mathrm{T}} & L_u \end{bmatrix} \begin{bmatrix} p_s^k \\ p_u^k \end{bmatrix} \quad (1)$$

Where $p^k = [p_1^k, \dots, p_N^k]$ denote the probability vector of the superpixels for the background label ($k = 1$) and foreground label (lesions) ($k = 2$). The probability was based on the FCN results and then averaged to the superpixels. The seeded superpixel is p_s^k and p_u^k is the unseeded superpixel. The foreground and background seeds were set to:

$$\begin{cases} p_s^k < mean(p^k), k = 1 \\ p_s^k > \frac{mean(p^k) + \max(p^k)}{2}, k = 2 \end{cases} \quad (2)$$

B is a partitioned matrix and L is a Laplacian matrix. While minimizing the Dirichlet function, we can find the critical point of the unseeded superpixels. We used 5 different scales

(from 200 to 1000 superpixels per slice with an increment of 200) of superpixels to address the issue of lesions of varying sizes. We used linear spectral clustering (LSC) to generate superpixels; LSC allows the superpixels to account for image-wide properties e.g., intensity variability [13].

We assumed for the multi-scale integration that the results produced on different scales were complementary to each other. Therefore, we applied cellular automata (CA) [14-16] with the synchronous updating rule to produce the final map (Figure 4g) and to take advantage of the superiority of each result on different scales. We treat the pixels that having the same coordinates on different scales of SRW results as the neighborhood pixels and we assume that these neighborhood pixels will have the same influence to determine the pixels' future status and this can be calculated as:

$$A(sc, t+1) = A(sc, t) + \sum_{si \in M/sc} \gamma \cdot sign(A(si, t) - \varphi_{si} \cdot [1, \dots, 1]^{T}) \quad (3)$$

Where $A(sc, t)$ represents the probability at scale sc at the time stamp t; M represents all different scales; γ is a constant weight and was set to 0.15 empirically; φ_{si} represents a threshold value that could binarize the original probability and the Otsu threshold was used for simplicity. When $t = 0$, we used the five different scales of SRW results. We used $t = 5$ iterations for convergence and averaged the different scales of CA to produce the final probability map.

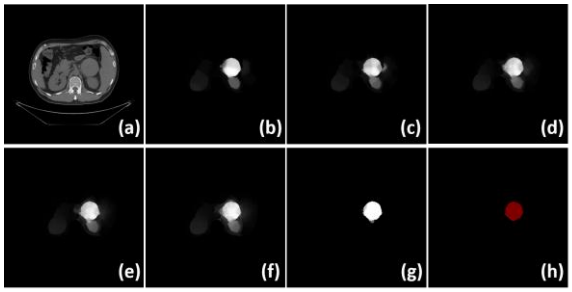

Figure 4. Multi-scale cellular automata based integration: (a) input image; (b-f) are the superpixel based RW probabilistic maps on different scales; (g) final integrated segmentation result, and (h) ground truth of the input image.

3. RESULTS AND DISCUSSIONS

3.1. Materials and Experimental Setup

Our dataset consisted of 38 low contrast whole-body CT studies with 38 adrenal lesion patients provided by the Department of Endocrinology and Metabolism, Ruijin Hospital affiliated to the Shanghai Jiao Tong University, Shanghai, China. All studies were acquired on a GE Discovery STE PET-CT scanner (General Electric Healthcare, Milwaukee, WI, USA) with a CT resolution of 512×512 pixels at 0.98mm^2 and slice thickness of 3.75mm. An experienced operator manually annotated each adrenal lesion for the ground truth; then all the annotated lesions were checked by a senior expert. We carried out our experiments on transaxial slices and excluded slices that did not contain lesions. We analyzed 553 slices and we randomly separated the CT slices into 2 roughly equally sized groups. We used the first group as the training and tested on the second group and then reversed the roles of the groups. We ensured that no patient CT slices were in both training and test groups.

We compared our approach to the state-of-the-art methods. The comparison methods included: (i) A SP-CNN-SVM technique. It is a superpixel based detection technique, which is similar to the works proposed in [17, 18]. We replaced the texture features with a new features set derived from a domain transferred deep CNN features (VGGNet) and classified via a support vector machine (SVM). Transfer learned CNN features have consistently shown benefits in the medical domain [9]; (ii) A SP-CNN-RF technique. This technique is similar to (i), we replaced SVM with a random forest (RF) classifier; (iii) The FCN technique [7]. This is a fully convolutional network and fine-tuned with a stride size of 8 (FCN-8s); (iv) A FCN-RW technique where we used FCN-8s with a superpixel based random walk. The evaluation was based on the overlap ratio between the detected lesion and the ground truth [5]. A detected tumor with >40% overlap with the ground truth was considered as true positive (TP). An additional detected lesion was considered as false positive (FP). We regarded an annotated lesion that was not detected, or an overlap, smaller than 40%, between the detected tumor and the annotated region as false negative (FN). We measured the overall precision (P), recall (R), f-score (F) and the precision, recall performance at different overlap ratios for all methods.

3.2. Results

Table 1: All detection results; numbers in **bold** indicate the best results

	R (%)	P (%)	F (%)	TP	FP	FN
SP-CNN-SVM	51.80	18.32	27.07	288	1284	268
SP-CNN-RF	52.63	27.46	36.09	290	766	261
FCN	46.62	53.58	49.86	262	227	300
FCN-RW	60.68	56.74	58.64	341	260	221
Our approach	**76.29**	**69.82**	**72.91**	**428**	**185**	**133**

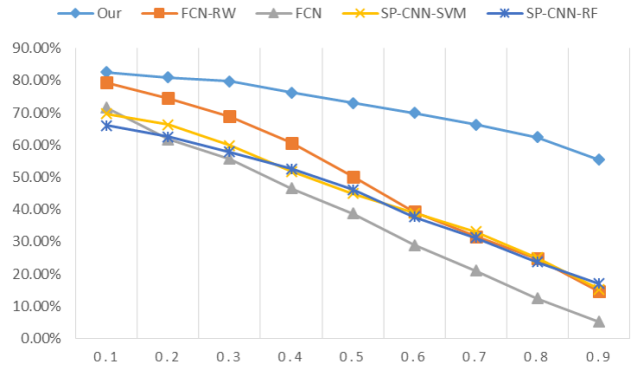

Figure 5. Recall performance (y-axis) at different overlap ratios (x-axis) for all approaches.

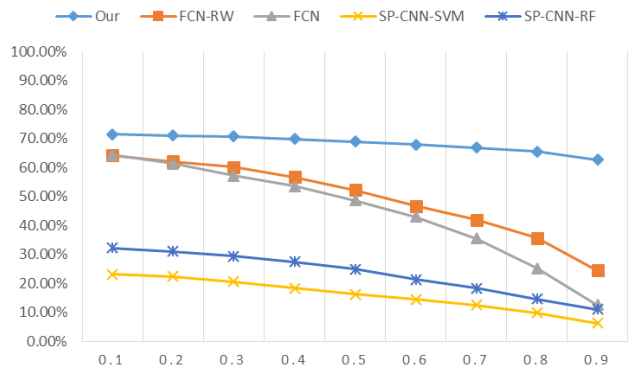

Figure 6. Precision performance (y-axis) at different overlap ratios (x-axis) for all approaches.

The detection results are shown in Table 1. Our approach had the best performance across the different techniques. Recall and precision in regard to different overlap ratios is shown in Figure 5 and Figure 6 and indicate that our approach had the best results.

3.3. Discussion

Our approach had the overall best performance when compared to other current 'state-of-the-art' approaches. Although SP-CNN-SVM and SP-CNN-RF had higher recall values when compared with FCN (see Table 1 and Figure 5), their precision values were lower. This can be explained by the inability of SP-CNN-SVM and SP-CNN-RF to differentiate tumor superpixels from surrounding tissues and resulted in false positive lesion detection. The FCN-RW approach showed good detection results and we suggest that this relates to superpixel-based random walk being able to leverage FCNs to refine lesion boundaries. Our proposed approach further improved upon the FCN-RW in recall with a 16% increase. We suggest that this is related to MI-SRW encouraging label agreement among neighboring pixels and constraining the spatial and appearance consistency.

4. CONCLUSION

We propose a new approach for the detection of adrenal lesions on low contrast CT images by using a FCN and multi-scale integration. Our preliminary results on 38 adrenal lesion studies indicate that our approach was the best performed when compared to existing approaches. In the future, we will apply our approach to a larger clinical dataset.

5. ACKNOWLEDGEMENT

This research was funded by Australian Research Council (ARC) grants.

REFERENCES

[1] M. Yun, W. Kim, N. Alnafisi, L. Lacorte, S. Jang, and A. Alavi, "18F-FDG PET in characterizing adrenal lesions detected on CT or MRI," Journal of Nuclear Medicine, vol. 42, pp. 1795-1799, 2001.

[2] M. A. Blake, M. K. Kalra, A. T. Sweeney, B. C. Lucey, M. M. Maher, D. V. Sahani, et al., "Distinguishing Benign from Malignant Adrenal Masses: Multi–Detector Row CT Protocol with 10-Minute Delay 1," Radiology, vol. 238, pp. 578-585, 2006.

[3] Sahdev, J. Willatt, I. R. Francis, and R. H. Reznek, "The indeterminate adrenal lesion," Cancer Imaging, vol. 10, pp. 102-113, 2010.

[4] G. W. Boland, M. A. Blake, P. F. Hahn, and W. W. Mayo-Smith, "Incidental adrenal lesions: principles, techniques, and algorithms for imaging characterization 1," Radiology, vol. 249, pp. 756-775, 2008.

[5] Y. Song, W. Cai, J. Kim, and D. D. Feng, "A Multistage Discriminative Model for Tumor and Lymph Node Detection in Thoracic Images," IEEE Transactions on Medical Imaging, vol. 31, pp. 1061-1075, 2012.

[6] D. Pescia, N. Paragios, and S. Chemouny, "Automatic detection of liver tumors," in IEEE International Symposium on Biomedical Imaging (ISBI), pp. 672-675, 2008.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440, 2015.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[9] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, et al., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," IEEE Transactions on Medical Imaging, 35.5, 1285-1298. 2016.

[10] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," arXiv preprint arXiv:1405.3531, 2014.

[11] Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in ACM international conference on Multimedia, pp. 689-692, 2015.

[12] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. D. Feng, "Robust Saliency Detection via Regularized Random Walks Ranking," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2710-2717, 2015.

[13] Z. Li and J. Chen, "Superpixel Segmentation using Linear Spectral Clustering," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1356-1363, 2015.

[14] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency Detection via Cellular Automata," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 110-119, 2015.

[15] L. Bi, J. Kim, E. Ahn, D. Feng, and M. Fulham, "Automated skin lesion segmentation via image-wise supervised learning and multi-scale superpixel based cellular automata," in IEEE International Symposium on Biomedical Imaging (ISBI), 2016, pp. 1059-1062.

[16] L. Bi, J. Kim, L. Wen, A. Kumar, M. Fulham, and D. D. Feng, "Cellular automata and anisotropic diffusion filter based interactive tumor segmentation for positron emission tomography," in IEEE Annual International Conference of Engineering in Medicine and Biology Society (EMBC), pp. 5453-5456, 2013.

[17] N. Yu, J. Wu, S. P. Weinstein, B. Gaonkar, B. M. Keller, A. B. Ashraf, et al., "A superpixel-based framework for automatic tumor segmentation on breast DCE-MRI," in SPIE Medical Imaging, pp. 94140O-94140O-7, 2015.

[18] L. Bi, J. Kim, A. Kumar, L. Wen, D. Feng, and M. Fulham, "Automatic detection and classification of regions of FDG uptake in whole-body PET-CT lymphoma studies," Computerized Medical Imaging and Graphics (CMIG), 2016