Neurocomputing 000 (2017) 1–10

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations

Avi Ben-Cohen[a,*], Eyal Klang[b], Ariel Kerpel[b], Eli Konen[b], Michal Marianne Amitai[b,1], Hayit Greenspan[a,1]

[a] Faculty of Engineering, Department of Biomedical Engineering, Medical Image Processing Laboratory, Tel Aviv University, Tel Aviv 69978, Israel
[b] Sheba Medical Center, Diagnostic Imaging Department, Abdominal Imaging Unit, affiliated to Sackler School of Medicine, Tel Aviv University, Tel Hashomer 52621, Israel

ARTICLE INFO

Article history:
Received 9 February 2017
Revised 15 July 2017
Accepted 1 October 2017
Available online xxx

Communicated by Dr. Bo Du

Keywords:
Deep learning
Liver lesions
Metastases
Detection
CT

ABSTRACT

In this work we focus on liver metastases detection in computed tomography (CT) examinations, using both a global context with a fully convolutional network (FCN), and a local patch level analysis with superpixel sparse based classification. The task of detecting metastases in the liver, in particular the small metastases, is important for early detection of liver cancer. Using a combined global and local approach, we present a system that can enhance detection capabilities. Our data contains a development set with CT examinations from 20 patients with a total of 68 lesions and a testing set with CT examinations from 14 patients with overall 55 lesions, out of which 35% were considered small lesions (longest diameter ≤ 1.5 cm). Experiments using 3-fold cross-validation resulted in a true positive rate of 94.6% with 2.9 false positives per case. These results are clinically promising, and should lead to better detection capabilities, including of small lesions, which is critical in cancer diagnosis.

1. Introduction

Liver cancer is among the most frequent types of cancerous diseases, and was responsible for the deaths of 745,000 patients worldwide in 2012 alone [1]. CT is one of the most common modalities used for detection, diagnosis and follow-up of liver lesions, and specifically metastases [2]. The images are acquired before and after intravenous injection of a contrast agent with optimal detection of lesions in the portal phase (60–80 s post injection) images. Current radiological practice is to manually analyze the liver. This is a time-consuming task requiring the radiologist to search through a 3D CT scan which may include hundreds of slices and multiple lesions. For early detection purposes, it is of particular importance to detect what are known as "too small to characterize" lesions. This requires additional radiologist time and focus, and is considered a difficult challenge.

Automated computerized analysis tools are starting to emerge to assist clinicians in the detection, segmentation, and characterization of these lesions. There are a number of difficulties associated with automatic liver lesion analysis that make it a challenging task. These include: differences in the contrast enhancement behavior of liver lesions and the parenchyma, the low image contrast between the tissues because of individual differences in perfusion and scan time, and the considerable variation in lesion shape, texture, and size from patient to patient, etc. The lesion *segmentation* task has attracted a great deal of attention in recent years, including the MICCAI 2008 Grand Challenge [3], and the upcoming ISBI 2017 challenge [4]. The objective is to find a segmentation mask similar to the one manually circumscribed by experts. In this task, a segmentation result can miss one or two small lesions and still get a high score since the larger lesions were correctly segmented in the same case.

This article focuses on the liver lesion *detection* task. In the detection task, the objective is to find *all lesions*. In clinical practice, finding small lesions can support the radiologist's diagnosis to an even greater extent than detecting large obvious lesions. Therefore, in evaluating automated system performance, missing a small lesion reduces the score in the same way a large lesion does. To achieve this goal the simplest solution apparently is to define all detected lesions as candidates. The challenge is to maintain the high detection rate with a reasonable false positive rate. Thus here, we concentrated on detecting very small lesions, while eliminating the majority of the false positives. Several recent works that have started to approach this detection task can be found in [5,6].

* Corresponding author.
E-mail address: avibenc@mail.tau.ac.il (A. Ben-Cohen).
[1] These authors contributed equally.

https://doi.org/10.1016/j.neucom.2017.10.001
0925-2312/

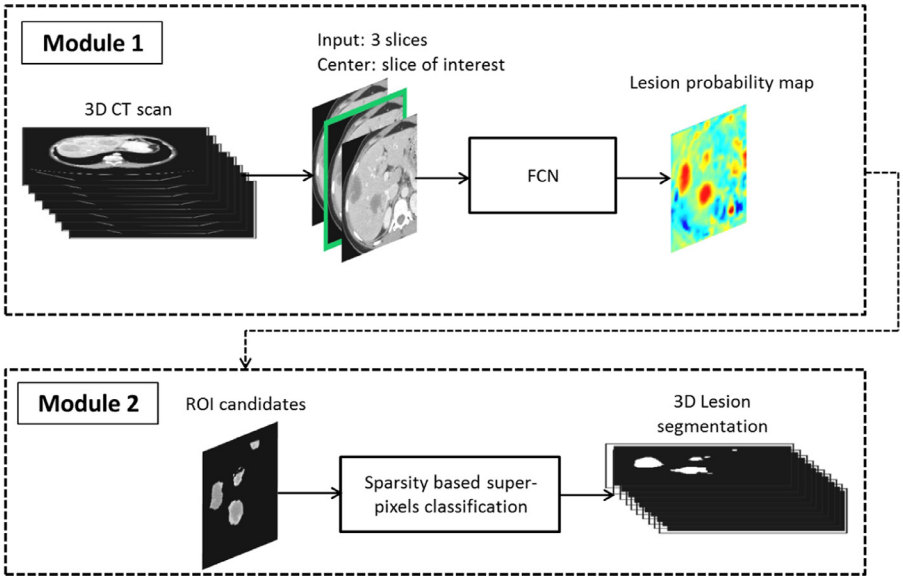


Fig. 1. The proposed lesion detection framework.

In recent years, deep learning has become a dominant research topic in numerous fields. In computer vision one common architecture is the Convolutional Neural Network (CNN), which has been shown to exhibit outstanding performance in visual object recognition and image classification in general imagery as well as more recently in the medical domain [7]. CNNs, introduced by LeCun et al. [8], are fully data-driven and can retrieve hierarchical features automatically by building high-level features from low-level ones, thus obviating the need to manually customize hand-crafted features. CNNs have been used for detection in several medical applications including pulmonary nodules [9], sclerotic metastases, lymph node and colonic polyp [10] and liver lesions [11]. In these works, the CNN was trained using patches taken from relevant regions of interest (ROIs).

A fully convolutional network (FCN) [12] presents an alternative approach to image segmentation, where the image is analyzed globally, instead of using localized patches. Thus, there is no need to select representative patches, eliminate redundant calculations where patches overlap, or scale up more efficiently with image resolution. Moreover, there can be a fusion of different scales by adding links that combine the final prediction layer with lower layers with finer strides. The FCN can take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. Unlike patch based methods, the loss function using this architecture is computed over the entire image segmentation result. The fully convolutional architecture was recently used for liver and lesion segmentation in CT examinations. Christ et al. [13] used cascaded fully convolutional neural networks and dense 3D conditional random fields to segment the liver, and achieved very high segmentation scores. Dou et al. [14] used a deeply supervised 3D fully convolutional network with additional supervision in some hidden layers to counteract the adverse effects of vanishing gradients for liver segmentation. They used a conditional random field model for the contour refinement and achieved low errors in the segmentation task. In previous work [15] we explored the fully convolutional architecture for the task of liver segmentation and reported the initial results for liver metastasis detection. We showed that the FCN outperformed a patch based CNN.

In this paper our goal is to describe an automated system that can analyze full CT examinations, and detect all lesions with high accuracy, specifically by exploring detection performance of small lesions. To overcome the hurdles inherent to both the localized, patch-based approach as well as a global FCN approach, we propose a system that combines global context via an FCN, along with local patch level analysis using superpixel sparse based classification. The combination of global and local analysis for false positive reduction has recently been shown to be a promising approach in a range of medical applications [9,10,16]. We evaluate the system on a dataset composed of varying size lesions including ones that are too small to characterize. To the best of our knowledge, this is the first work to combine the FCN with online sparsity based learning for this task.

The method is presented in Section 2. The experimental results are described in Section 3. The experiments include comparison of various algorithmic solutions to the task along with an evaluation of our method with regard to lesion size. We conclude this paper with a discussion in Section 4.

2. Methodology

Our liver lesion detection framework is made up of two main modules, as shown in Fig. 1. The first module is an FCN. Following the radiology procedure, we input axial slices to search for the lesions. Since lesions are present in several sequential slices, three slices are provided to the FCN: the target slice in the center and two adjacent slices above and below. Using an FCN – based analysis, this module outputs a lesion probability map. Based on the high-probability candidate lesion regions from the first module, the second module follows with localized patch level analysis using superpixel sparse based classification. This module's objective is to classify each localized superpixel as a lesion or not. Thus it provides a fine-tuning step, with the objective of increasing sensitivity to lesions while removing false positives (FPs). In the following subsections we provide details about the FCN architecture used, the sparsity based classification scheme and their combination in our final framework.

2.1. Global analysis via a fully convolutional network architecture

The FCN architecture that we implemented in this work is based on the FCN-8s net, using the VGG-16 layer net, as in [12,17]. The FCN-8s architecture learns to combine coarse, high layer

Please cite this article as: A. Ben-Cohen et al., Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.10.001

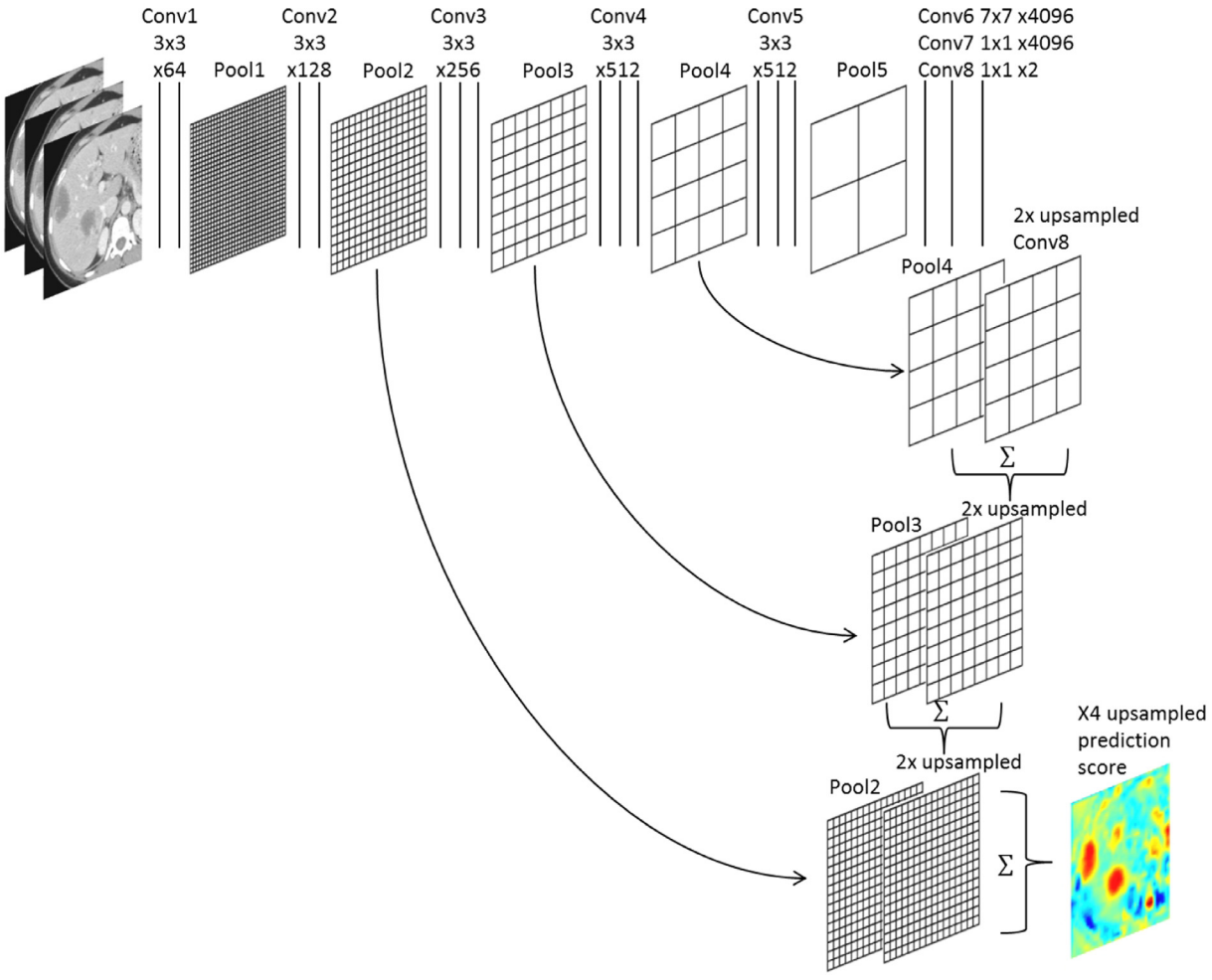


Fig. 2. FCN-4s network architecture. Each convolution layer is illustrated by a straight line with the receptive field size and number of channels denoted above. The ReLU activation function and drop-out are not shown for brevity.

information with fine, low layer information as follows. The VGG-16 net is decapitated by discarding the final classifier layer, and converting all fully connected layers to convolutions. We appended a 1×1 convolution with channel dimension 2 to predict scores for lesion or liver at each of the coarse output locations, followed by a deconvolution layer to upsample the coarse outputs to pixel-dense outputs. The upsampling was performed in-network for end-to-end learning by backpropagation from the pixelwise loss. In earlier work we showed that for liver lesion analysis, additional lower level linking is advantageous [15]. This is done by linking the second max-pool layer in a similar way to the linking of the third and fourth max-pool layers. The final FCN network architecture, termed FCN-4s net, is presented in Fig. 2. The fusion steps within the FCN architecture were shown to support the analysis of different lesion sizes. The output of the network is a lesion probability map.

2.1.1. Input to the network

To achieve fixed in-slice resolution across all patient data, as well as fixed slice spacing in the network input, we resized (using linear interpolation) to obtain a fixed pixel spacing of 0.71 mm and a fixed slice spacing (between the adjacent slices) of 1 mm.

When using deep networks, it is common to augment the input data. This step is essential to teach the network the desired invariance and robustness properties when limited data are available. In this work, we focus on liver metastases, which varied considerably in size. To reflect this in the training data, we applied different scale transformations to the available training images in each batch. For each input image in each batch, two numerical values between 0.9 and 1.1 were randomly selected (uniform distribution). Two corresponding scale transformations were applied to the image (3 adjacent slices), using nearest-neighbor for resampling, to generate two training images.

2.1.2. Training of the network

Input images along with the corresponding liver and lesion segmentation masks provided by the human expert were used to train the network. Examples of masks are shown in Fig. 5. We used the stochastic gradient descent implementation of MatConvNet [18] with GPU acceleration (nvidia GTX 1080). Regions surrounding the liver, including the different organs and tissues, were ignored during training. The softmax log-loss function was computed pixel-wise with different weights for each class of pixels as in Eq. (1):

$$L = -\sum_{ij} w_{ij} \left(x_{ijc} - \log \sum_{d=1}^{D} \exp(x_{ijd}) \right) \tag{1}$$

where $c \in [1...D]$ is the ground-truth class, x is the prediction score matrix (before softmax), and w is the per-pixel weight matrix. Most of the pixels in each image belonged to the liver, and a much smaller number belonged to the lesions. The learning process was therefore balanced using fixed weights that were inversely

Please cite this article as: A. Ben-Cohen et al., Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.10.001

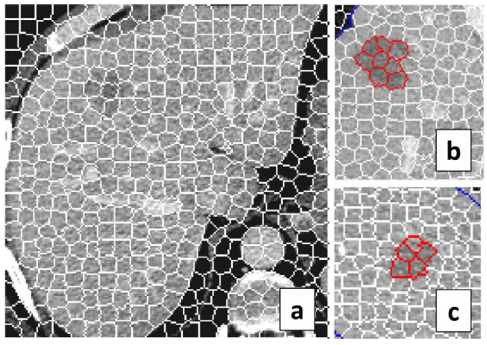


Fig. 3. SLIC superpixels on liver CT image. (a) An entire liver divided into superpixels (in white); (b, c) two lesion examples (zoom-in) divided into superpixels (in red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

proportional to the population ratios. The learning rate was chosen to be 0.0005 for the first 30 epochs and 0.0001 for the last 40 epochs (total of 70 epochs). The weight decay was chosen to be 0.0005 and the momentum parameter was 0.9.

2.2. Localized analysis via superpixel sparse based classification for false-positives reduction

In the second main module of the proposed system (Fig. 1), a more localized analysis is conducted. The input to this phase are lesion candidates (ROIs) from the FCN network output. For each candidate region, a superpixel map is generated. The rationale for using superpixels as well as the generation of the superpixel representation are described in detail below. For each superpixel, a representative feature vector is extracted. Using a pre-learned dictionary, a sparse-code representation is computed and classified. In the following subsections we describe the dictionary learning process and the dictionary fine-tuning step that served to produce the final classification map.

2.2.1. Super-pixel representation

Superpixels provide an efficient way to select patches of interest and avoid the redundancy of densely extracted patches. We divided each liver in each CT examination into a set of homogeneous superpixels. Each image was clustered into N superpixels using the well-known SLIC algorithm [19], where N is chosen so that each superpixel will be approximately of size 25 mm^2. The SLIC algorithm is known to preserve strong edges. As Fig. 3 illustrates, this ensures that the superpixels align to preserve the overall liver boundary and the lesion boundaries. As in our network training, non-liver superpixels were not included in the classification.

Each superpixel was represented by a 125 long feature vector. The features included histogram statistics and various textural features such as Gabor filters, local binary patterns (LBP) [20] and acutance [21] around the superpixel boundary. These features provide information about the superpixels' gray level distributions, edge behavior, texture patterns, and contrast around the superpixels' boundaries.

A superpixel was defined as a lesion if most of its pixels overlapped with the lesion segmentation mask provided by the expert.

2.2.2. Superpixel dictionary learning

To distinguish between lesion and non-lesion superpixels, we used a sparsity based classification method called Label consistent K-SVD (LC-KSVD) [22]. First, a reconstructive dictionary was built for initialization where the feature vectors extracted from the superpixels are fed into K-SVD to generate an initial dictionary for the lesion discriminative learning step [6].

The LC-KSVD method aims to leverage the supervised information of input signals or feature vectors to learn a reconstructive and discriminative dictionary. The dictionary items were chosen so that each one represented a subset of the training feature vectors ideally from a single class. Hence, each dictionary item d_k could be associated with a single label. Thus there was an explicit correspondence between dictionary items and the appropriate labels. This connection was created using a label consistency regularization term and a joint classification error in the objective function for learning a dictionary with more balanced reconstructive and discriminative power.

In the following we provide a short overview of this process. Let Y be a set of n-dimensional N feature vectors extracted from the superpixels; i.e., $Y = [y_1, \ldots, y_N] \in R^{n \times N}$. This method solves the following objective function:

$$\begin{aligned} \langle D, W, A, X \rangle = arg \min_{D,W,A,X} & \|Y - DX\|_2^2 \\ & + \alpha \|Q - AX\|_2^2 \\ & + \beta \|H - WX\|_2^2 \\ & s.t \quad \forall i \quad \|x_i\|_0 \le T \end{aligned} \tag{2}$$

where $\|Y - DX\|_2^2$ denotes the reconstruction error and $D = [d_1, \ldots, d_k] \in R^{n \times K}$ is the learned dictionary, $X = [x_1, \ldots, x_N] \in R^{K \times N}$ are the sparse codes of input vectors Y, and T is a sparsity constraint factor. The relative contribution between reconstruction and label consistency regularization is controlled by parameter α and $Q = [q_1, \ldots, q_N] \in R^{K \times N}$ are the discriminative sparse codes of input vectors Y for classification. $q_i = [q_i^1, \ldots, q_i^K]^t = [0, \ldots, 1, 1, \ldots, 0]^t \in R^K$ is a discriminative sparse code corresponding to an input vector y_i if the nonzero values of q_i occur at those indices where the input vector y_i and the dictionary item d_k share the same label. A is a linear transformation matrix, which transforms the original sparse codes x to be the most discriminative in sparse feature space R^K. The term $\|Q - AX\|_2^2$ represents the discriminative sparse code error, which enforces that the transformed sparse codes AX approximate the discriminative sparse codes Q. It forces the feature vectors from the same class to have very similar sparse representations. The term $\|H - WX\|_2^2$ represents the classification error. W denotes the linear classifier parameters ($f(x, W) = Wx$). $H = [h_1, \ldots, h_N] \in R^{m \times N}$ are the class labels of input signals Y . $h_i = [0, 0, \ldots, 1, \ldots, 0, 0]^t \in R^m$ is a label vector corresponding to an input feature vector y_i, where the nonzero position indicates the class of y_i. β is a scalar controlling the relative contribution of this term.

The jointly learned linear classifier W can be used to classify superpixel representation as lesion or non-lesion. We used $\beta = 2$ and $\alpha = 3$ with a dictionary size of 200 items and a sparsity constraint factor value of 20 in our experiments based on a development dataset described in our previous work [6]. The classifier performance using different sparsity constraint factors is presented in Section 3.3. The F1-score was used to choose the best performance; the precision and recall are also presented.

Fig. 4 illustrates the localized framework: (a) the training step using the LC-KSVD method; (b) a fine-tuning step using the FCN output; (c) final classification of the superpixels. Steps (b) and (c) are described next.

2.2.3. Dictionary personalized fine-tuning

In addition to the training phase, we personalized the dictionary and classifier for each test case using the knowledge obtained from the FCN. An online fine-tuning step (Fig. 4(b)) was conducted using a selection of superpixels from each test case, based on the corresponding FCN probability map. Two probability thresholds were defined: *high-threshold=0.98* and *low-threshold=0.2*. Superpixels with an average probability of *high-threshold* and above

Please cite this article as: A. Ben-Cohen et al., Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.10.001

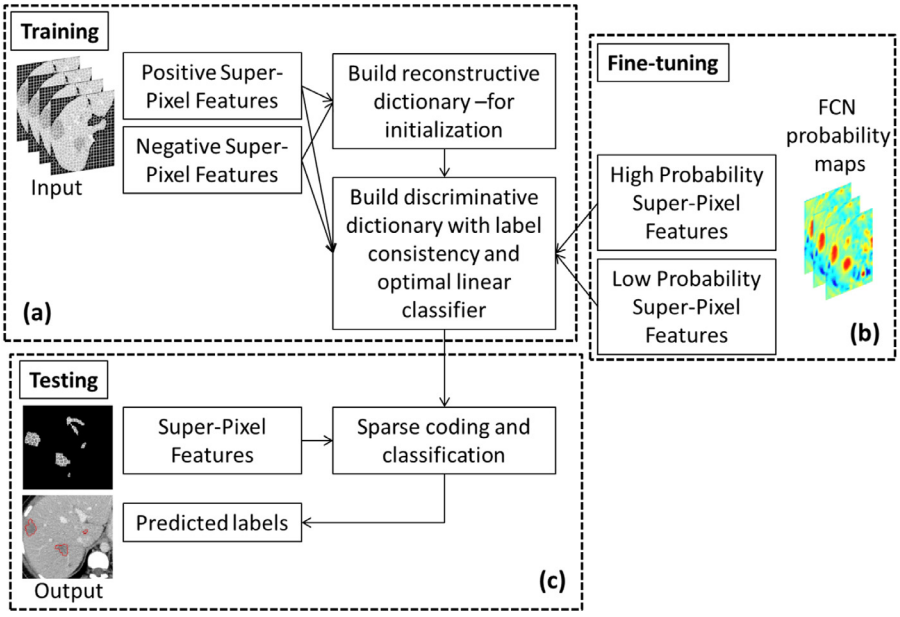


Fig. 4. The sparsity based framework: (a) training; (b) fine-tuning using the FCN probability maps from online learning; (c) testing.

were labeled "lesion" and those with an average probability of *low-threshold* and below were labeled "non-lesion". The sparsity based dictionary and the linear classifier parameters were fine-tuned using the FCN based labeling in cases where the number of lesion superpixels was at least half the dictionary size.

2.2.4. Final classification

Following the fine-tuning stage, all candidate superpixels from the first module were classified as lesion or non-lesion (Fig. 4(c)) using the jointly learned linear classifier, to provide a binary detection map for the input CT scan. Next we used a connected-component labeling algorithm to transform the binary detection map to a lesion detection map, where each connected-component represented a lesion detected by our system.

3. Experiments and results

To evaluate the performance of the system, we conducted several sets of experiments. These included evaluations of a development set followed by the results of an independent testing set. We focused on the lesion detection within the liver to evaluate the detection performance independent of automated liver segmentation capabilities. Thus, the majority of the experiments and evaluations were constrained to the manually circumscribed liver area. As a final experiment, we included our own liver segmentation algorithm in a fully automated liver segmentation and lesion detection framework to examine the initial results.

3.1. Data

The data used in the current work included CT scans from the Sheba Medical Center, obtained from 2009 to 2015. Different CT scanners were used with 0.71–1.17 mm pixel spacing and 1.25–5 mm slice thickness. Markings of the entire liver and hepatic metastases boundaries were conducted by a board certified radiologist (E.K. with eight years of experience) and a radiology resident (A.K. with one year of experience), in consensus. Fig. 5 shows examples of the segmentation masks provided by these radiologists.

Algorithms were developed using a development dataset including CT examinations from 20 patients with 68 lesions in total and were tested on a testing set that included CT examinations from 14 patients with a total of 55 lesions, out of which 35% were considered small lesions. In the development set, we obtained 1–2 marked slices per patient. In the testing set, we obtained marked slices in the entire CT volume (multi-slice). We defined a lesion as a small lesion if its longest diameter (LD) was less than 1.5 [cm] [23]. In our testing set we had 19 small lesions out of the 55 lesions (35%). The mean LD of all lesions in the development set was 39.2 ± 17.3 mm (7.1–55.3 mm) and 39.3 ± 28.4 mm (9.3–91.2 mm) for the lesions in the testing set. The data included various liver metastatic lesions derived from different primary cancers.

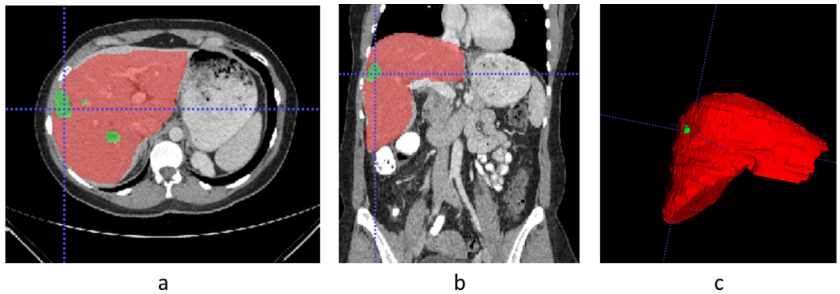


Fig. 5. Experts' segmentation masks. (a) Axial view; (b) coronal view; (c) 3D segmentation mask. In red: liver mask; in green: lesion mask. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: A. Ben-Cohen et al., Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.10.001

Table 1
Performance evaluation on the development set.

Method	TPR [%]	FPC
FCN-4s + sparsity based	88.2	0.53
FCN-4s	88.2	0.74
U-Net	83.8	3.0
Sparsity based	82.4	1.1
Patch-based CNN	85.3	1.9

Table 2
Performance evaluation on the testing set.

Method	TPR [%]	FPC
FCN-4s + sparsity based	94.6	2.9
FCN-4s	92.7	4.2
U-Net	92.7	7.6
Sparsity based	83.6	9.2
Patch-based CNN	78.2	9.7

3.2. Comparative system performance

Detection performance was assessed using the following two metrics: the true positive rate (TPR) defined as the total number of detected lesions divided by the total number of known lesions, and the false positive per case (FPC) defined as the total number of false detections divided by the number of patients. We defined a lesion as detected when a certain connected component (CC) in the system output overlapped with the radiologist's marked lesion mask. Note that due to the data in our study, We extracted a 2D CC in the development set whereas for the testing set, the CC and corresponding overlap measures were in 3D.

We compared the proposed system to five alternative analysis schemes: (1) a *Superpixel sparse dictionary learning and classification – sparsity based scheme*. In this scheme, the CT input is divided into superpixels which are directly input to the superpixel dictionary learning, and classified using superpixel sparse-based classification; a localized, patch-based analysis; (2) *Patch-based CNN* based on Li et al. [11], which used patches of 17×17 pixels which are fed to the CNN for classification into lesion/normal area. A CNN model with seven hidden layers was used, which included three convolutional layers, two max-pooling layers, a fully connected layer, ReLU and a softmax classifier; (3) The *U-Net* scheme. Here, the proposed FCN architecture is replaced by the well-known U-Net architecture [24] with no use of localized analysis; (4) FCN-4s: first module only with no use of localized analysis; and (5) *FCN-4s+ Sparsity Based scheme*: Our proposed system, combining the FCN global analysis with a sparsity-based superpixel analysis. A 3-fold cross validation was used (each group containing different patients) for both the development set and the testing set using the experts' defined masks as the gold standard. Each test fold included approximately 7 patients in the development set and 5 patients in the testing set. Average results are presented for all patients (20 in the development set and 14 in the testing set) using cross validation.

Table 1 presents the results for the development set. Using the sparsity-based superpixel analysis scheme on its own generated the lowest TPR, at a reasonable FPC. Comparison of the two global analysis schemes that we compare (FCN-4s and U-Net) indicated that the FCN-4s had a higher TPR with a lower FPC. Combining the global FCN-4s with the sparsity-based superpixel analysis maintained a high detection performance of more than 88%, with a reduction in the FPC.

Table 2 lists the results obtained using the testing set. A trend similar to performance on the development set was observed. The FCN-4s with the sparsity based false positive reduction had the highest detection performance compared to other methods with a TPR of 94.6% and a significant improvement (P-value < 0.001) in the number of false-positives with a 2.9 FPC.

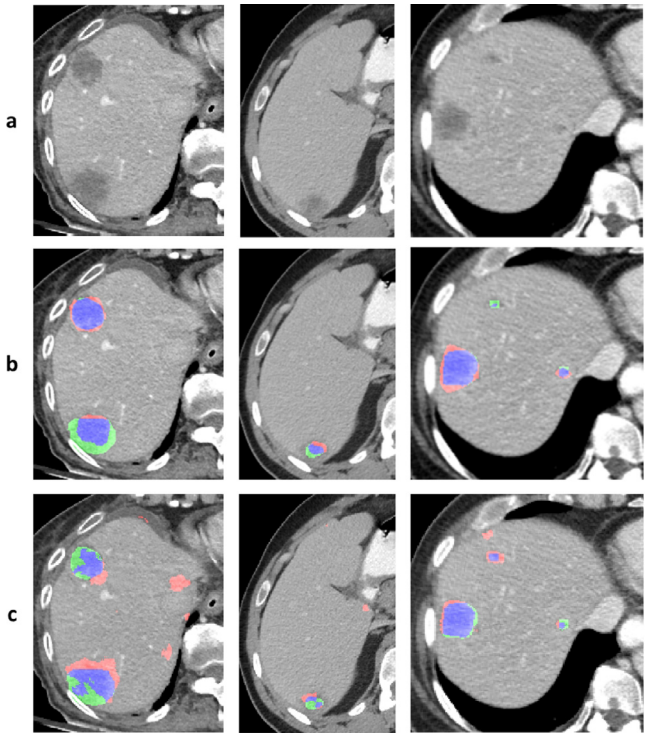


Fig. 6. Lesion detection sample results. (a) Input-CT slices (cropped); (b) detection candidates using our method; (c) from left to right: detection candidates using the patch-based CNN, sparsity based, and U-Net method. For each connected component the following conventions were used: in red – false positive pixels, in green – false negative pixels, in blue – true positive pixels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Detection performance for small and normal sized lesions on the testing set.

Method	TPR [%] All	FPC All	TPR [%] Normal	TPR [%] Small
FCN-4s + sparsity	94.6	2.9	94.4	94.7
FCN-4s	92.7	4.2	91.7	94.7
U-Net	92.7	7.6	91.7	94.7
Sparsity based	83.6	9.2	91.7	68.4
Patch-based CNN	78.2	9.7	88.9	63.2

Fig. 6 shows sample results using our method: (a) the CT input slices and (b) the output results. The results show the per-pixel labels as true positive (TP), false positive (FP), and false negative (FN). Using this visualization, blobs including blue regions (true positive) represent correctly detected lesions; blobs that only include only a red region are defined as false positives, and blobs that only include a green region are defined as false negatives.

Table 3 depicts the results for small sized lesions with a LD of less than 1.5 [cm] and normal sized lesions with a LD above 1.5 [cm]. When adding the small candidate detection, the best results were obtained using the FCN-4s with the sparsity based false positives reduction which yielded 2.9 FPC and 94.6% TPR. A TPR of 94.4% was obtained for the normal sized lesions and 94.7% for the small lesions. The sparsity based false positives reduction increased the TPR, which occurred in cases where two correctly detected candidates were connected by a false classification of some voxels. In these cases we obtained one connected component instead of two. The reduction of the connecting voxels added another correct detection of a lesion and increased the TPR.

Please cite this article as: A. Ben-Cohen et al., Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.10.001

Table 4
Detection performance of our method with and without the dictionary online fine-tuning on the testing set.

Method	TPR [%]	FPC
Fine-tuning	94.6	2.9
No fine-tuning	92.7	3.15

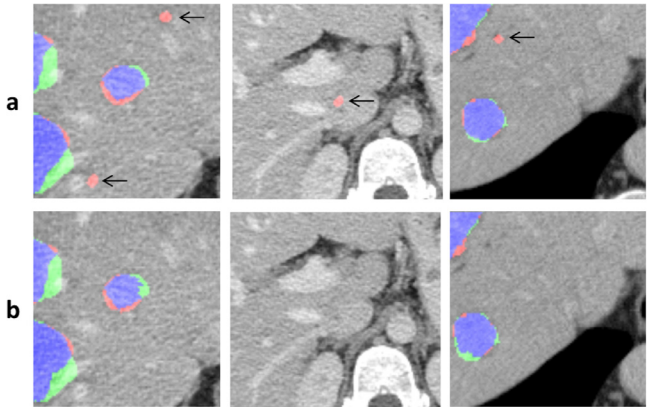


Fig. 7. Lesion detection sample results for (a) the FCN and (b) the FCN with the sparsity based FP reduction. The black arrows point to the FP candidates. In red – false positive pixels; in green – false negative pixels; in blue – true positive pixels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In addition, we analyzed the average per-slice runtime of each method: FCN-4s – 0.178s; FCN-4s + sparsity – 0.489s; U-Net – 0.578s; Sparsity based – 2.023s; Patch-based CNN – 5.321s.

3.3. System evaluation

In the next set of experiments, we evaluate several components of the system. These demonstrate the contribution of the system fine-tuning step. In Table 4 the results with and without the online dictionary fine-tuning are presented. The online fine-tuning improved the TPR slightly and reduced the FPC.

Fig. 7 shows sample results with and without the sparsity based FP reduction. In (a) FPs (red blobs marked with arrows). These were removed in (b) using the proposed system.

One key parameter that must be defined in the system is the probability threshold, *th*, which defines the set of candidate regions extracted from the FCN probability maps. To demonstrate the system behavior, the algorithm was executed with different probability thresholds. Figs. 8 and 9 show the free-response receiver operating characteristic (FROC) [25] with FPC and TPR for various *th* values with and without the sparsity based FP reduction for the test set. Unlike the classic FROC curve where the TPR increases for decreasing *th*, here there was a moderate decrease for *th* below 0.65. The TPR can decrease when two candidates merge into one candidate even though there are two lesions. Comparing the FROC curves in Figs. 8 and 9, the improvement in the FPC using the sparsity based FP reduction can be seen for each one of the tested thresholds. We used $th = 0.95$ in our experiments.

In addition, we tested the sparsity based super-pixel classification using different sparsity constraint factors, *T*, on the development set. Fig. 10 shows the Precision, Recall, and F1 score using different *T* values. The F1 score was used to choose the final value (T=20).

3.4. Automatic liver segmentation and lesion detection

In the final experiment, we tested a clinically viable scenario, in which the automated system needs to first localize the liver, following which it can detect the lesions within it. For this initial experiment, we used our previously developed liver segmentation network which is based on the FCN-8s architecture (for details see [15]). Examples of automatic liver segmentation can be seen in Fig. 11. Following the liver segmentation output, we used our lesion detection framework to automatically detect the lesions in the liver segmentation results.

We achieved a TPR of 90.9% and an FPC of 3.0 using the fully automatic algorithm. These results are comparable to the detection results achieved using the manual segmentation of the liver with a lower TPR and slightly higher FPC.

4. Discussion

In this work, we developed a computer-aided detection system to detect liver lesions on CT images using an FCN with a sparsity based online fine-tuning for false positive reduction. The FCN-4s architecture was shown to successfully detect lesions in both the development as well as the testing sets, as shown in Tables 1 and 2. The sparsity based classification scheme was shown to significantly reduce the FPC. We note a difference in the range of the results in the two sets. Due to the difference in the data settings. The data we obtained in the testing set had multi-slice lesion segmentation masks with only 1–2 slices marked in the development set. Given the additional data per patient, most lesions appeared on several slices in the test set, and thus had a higher likelihood of being detected. We therefore expected an overall increase in the TPR as well as in the FPC. Both these trends were present in the results.

The experimental results listed in Table 3 demonstrate that the FCN-4s method with the sparsity based fine-tuning is robust to lesion size, detecting 94.7% of the small lesions. Fig. 6 shows the detection results for several patients with different lesion sizes. Although all the lesions were detected, in some cases the overlap with the experts' segmentation mask was smaller than in others. However, as detailed below, we believe that the output of the detection system can serve as an initial lesion segmentation mask.

We compared our FCN architecture to the U-Net architecture in Tables 1 and 2. In both the development and testing sets the U-Net had a higher FPC than the FCN-4s. Thus, whereas the U-Net architecture exhibited superior segmentation performance in several medical applications, in the detection task the FCN-4s showed superiority. Using a FROC curve, we analyzed the TPR and FPC for various probability thresholds. Fig. 8 shows that for lower thresholds the FPC was increased as expected; however, a small decrease in TPR was observed since some proximal detected lesions were merged into one detected lesion by our method. Analyzing the FROC curve of the FCN without the sparsity based FP reduction, as shown in Fig. 9, resulted in a much higher FPC for all the thresholds tested.

Most of the experiments in our study were designed for detection performance evaluation, and concentrated on a manually circumscribed liver region. As a final experiment, we evaluated the proposed method using an automated liver segmentation tool. In our initial experiment, we achieved encouraging results of 90.9% TPR with 3.0 FPC with this fully automated approach.

In this work we focused on lesion detection. We did not include lesion segmentation quantification measures. To further enhance the system to provide accurate segmentation, these detection results can be used as an initial mask for the lesions, after which additional segmentation algorithms can be implemented for expansion and refinement. Note that no significant pre-processing or post-processing was implemented in the suggested method. Adding these steps could increase lesion detection accuracy and enable more accurate segmentation as well.

Please cite this article as: A. Ben-Cohen et al., Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.10.001

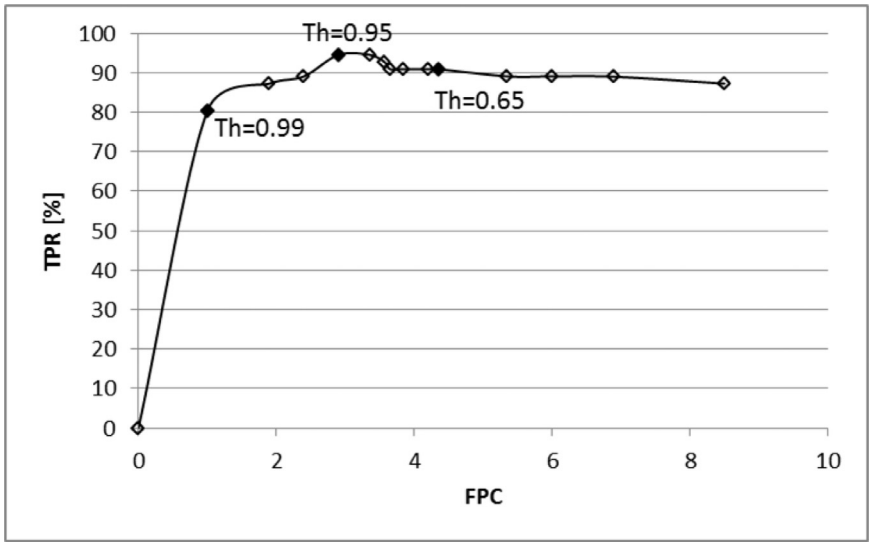


Fig. 8. FROC curve of lesion detection using FCN with sparsity based FP reduction on the testing set.

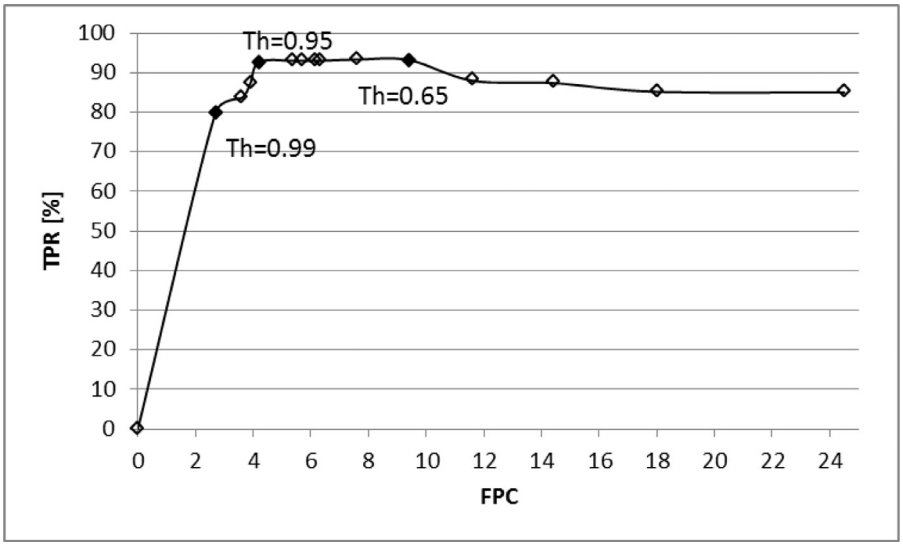


Fig. 9. FROC curve of lesion detection using FCN only on the testing set.

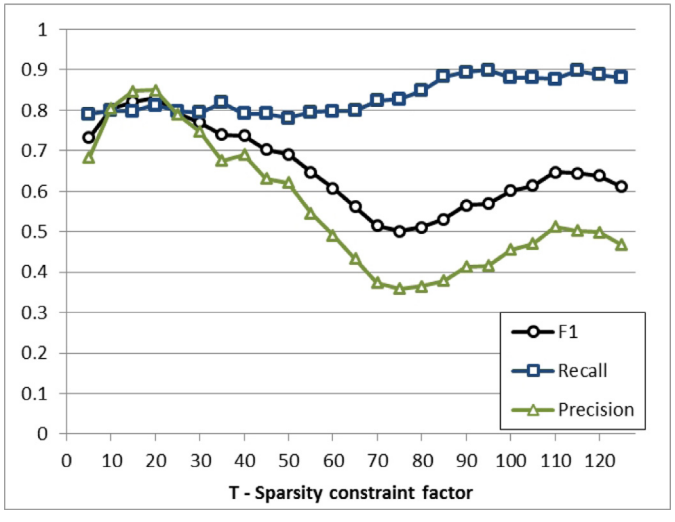


Fig. 10. F1, recall, and precision for superpixel classification using LC-KSVD on a development set for different sparsity constraint factors.

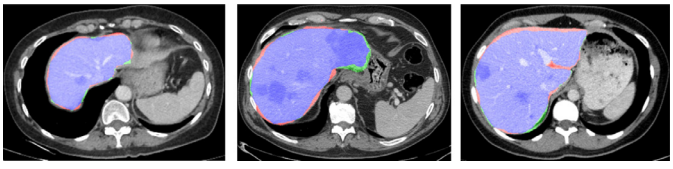


Fig. 11. Liver segmentation sample results. In red – false positives; in green – false negatives; in blue – true positives. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The FCN-4s in our system uses 3 adjacent slices as input. In examinations with large slice spacing, the images were interpolated to have a fixed slice spacing of 1 mm. For very small lesions, this means that a lesion will be present on a single slice alone which results in a lower TPR. It is clear that higher resolution in acquisitions will increase the overall detection rate. Moreover, using the automatic liver segmentation, a slight decrease in TPR was observed. A more accurate liver segmentation could improve the results of our fully automatic algorithm. The data augmentation in our training only included scale transformations. Additional augmentation techniques such as gray level transformations and deformations could make the system more robust to these changes.

Please cite this article as: A. Ben-Cohen et al., Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.10.001

Thus overall, we showed a fully automated convolutional network with sparsity based false positive reduction for the detection of liver metastases in CT examinations. The results indicate that the FCN-4s with data augmentation, addition of neighboring slices, appropriate class weights and false positive reduction using a sparsity based approach provides strong and robust results. In future work we plan to conduct clinical studies to evaluate the extent to which an automated system such as the system proposed here can facilitate the radiologists' workflow.

Disclosure of conflict of interest

The authors have no relevant conflicts of interest to disclose.

Acknowledgments

Part of this work was funded by the INTEL Collaborative Research Institute for Computational Intelligence (ICRI-CI).

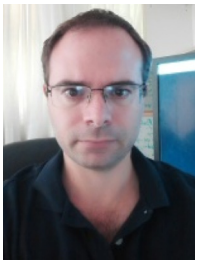
This research was supported by the Israel Science Foundation (Grant no. 1918/16).

Avi Ben-Cohen's scholarship was funded by the Buchmann Scholarships Fund.

References

[1] WHO, The World Health Report – World Health Organization, WHO, 2014.
[2] K.D. Hopper, K. Singapuri, A. Finkel, Body CT and oncologic imaging, Radiology 215 (1) (2000) 27–40, doi:10.1148/radiology.215.1.r00ap1727.
[3] X. Deng, G. Du, Editorial: 3D segmentation in the clinic: a grand challenge II – liver tumor segmentation, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention 2008, Springer, 2008.
[4] P.F. Christ, LiTS-Liver Tumor Segmentation Challenge, https://competitions.codalab.org/competitions/15595, 2017.
[5] L. Ruskó, Á. Perényi, Automated liver lesion detection in CT images based on multi-level geometric features, Int. J. Comput. Assist. Radiol. Surg. 9 (4) (2014) 577–593, doi:10.1007/s11548-013-0949-9.
[6] A. Ben-Cohen, E. Klang, M. Amitai, H. Greenspan, Sparsity-based liver metastases detection using learned dictionaries, in: Proceedings of 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), IEEE, 2016, pp. 1195–1198, doi:10.1109/ISBI.2016.7493480.
[7] H. Greenspan, B. van Ginneken, R.M. Summers, Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique, IEEE Trans. Med. Imaging 35 (5) (2016) 1153–1159, doi:10.1109/TMI.2016.2553401.
[8] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324, doi:10.1109/5.726791.
[9] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S.J. van Riel, M.M.W. Wille, M. Naqibullah, C.I. Sánchez, B. van Ginneken, Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks, IEEE Trans. Med. Imaging 35 (5) (2016) 1160–1169, doi:10.1109/TMI.2016.2536809.
[10] H.R. Holger, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, R.M. Summers, Improving Computer-aided Detection using Convolutional Neural Networks and Random View Aggregation, IEEE Trans. Med. Imaging (2015), doi:10.1109/TMI.2015.2482920.
[11] W. Li, F. Jia, Q. Hu, Automatic segmentation of liver tumor in CT images with deep convolutional neural networks, J. Comput. Commun. 3 (11) (2015) 146, doi:10.4236/jcc.2015.311023.
[12] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440, doi:10.1109/TPAMI.2016.2572683.
[13] P.F. Christ, M.E.A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. DAnastasi, et al., Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3d conditional random fields, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 415–423, doi:10.1007/978-3-319-46723-8_48.
[14] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, P.-A. Heng, 3d deeply supervised network for automatic liver segmentation from CT volumes, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 149–157, doi:10.1007/978-3-319-46723-8_18.
[15] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, H. Greenspan, Fully convolutional network for liver segmentation and lesions detection, in: Proceedings of International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Springer, 2016, pp. 77–85, doi:10.1007/978-3-319-46976-8_9.
[16] L. Lu, P. Devarakota, S. Vikal, D. Wu, Y. Zheng, M. Wolf, Computer aided diagnosis using multilevel image features on large-scale evaluation, in: Proceedings of International MICCAI Workshop on Medical Computer Vision, Springer, 2013, pp. 161–174, doi:10.1007/978-3-319-055530-5_16.
[17] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.
[18] A. Vedaldi, K. Lenc, Matconvnet: Convolutional neural networks for matlab, in: Proceedings of the 23rd ACM international conference on Multimedia, ACM, 2015, pp. 689–692, doi:10.1145/2733373.2807412.
[19] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282, doi:10.1109/TPAMI.2012.120.
[20] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognit. 29 (1) (1996) 51–59, doi:10.1016/0031-3203(95)00067-4.
[21] R.M. Rangayyan, N.M. El-Faramawy, J.L. Desautels, O.A. Alim, Measures of acutance and shape for classification of breast tumors, IEEE Trans. Med. Imaging 16 (6) (1997) 799–810, doi:10.1109/42.650876.
[22] Z. Jiang, Z. Lin, L.S. Davis, Learning a discriminative dictionary for sparse coding via label consistent K-SVD, in: Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 1697–1704, doi:10.1109/CVPR.2011.5995354.
[23] H.I. Khalil, S.A. Patterson, D.M. Panicek, Hepatic lesions deemed too small to characterize at CT: prevalence and importance in women with breast cancer, Radiology 235 (3) (2005) 872–878, doi:10.1148/radiol.2353041099.
[24] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241, doi:10.1007/978-3-319-24574-4_28.
[25] C.E. Metz, Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems, J. Am. Coll. Radiol. 3 (6) (2006) 413–422, doi:10.1016/j.jacr.2006.02.021.

Avi Ben-Cohen is a Ph.D. candidate at Tel-Aviv University under the supervision of prof. Hayit Greenspan. He holds a B.Sc. and an M.Sc. in biomedical engineering in the field of medical image processing. His current research interests include medical image processing, liver lesions analysis, computer-aided diagnosis and deep learning.



Eyal Klang received Diploma of Doctor in medicine from the Tel Aviv University, 2005. Completed radiology residency from 2009 to 2014 in the department of diagnostic imaging, the Chaim Sheba medical center at Tel Hashomer hospital and is a board certified radiologist. Since2014, he is working as a Staff Radiologist in the Department of Diagnostic Imaging, the Abdominal Imaging Unit, at the Chaim Sheba Medical Center at Tel Hashomer hospital. He lectures at the Sackler Faculy of Medicine, Tel Aviv University.

Ariel Kerpel received the B.Sc. and M.D. degrees from The Hebrew University of Jerusalem, Israel in 2011 and 2015, respectively. He is currently a Diagnostic Radiology Resident at The Chaim Sheba Medical Center, Tel HaShomer, Israel.

Eli Konen accomplished his medical studies in Sackler School of Medicine, Tel Aviv University and received his M.D. in 1987. Following a military career in the IDF, he accomplished his residency in Diagnostic Imaging in Sheba Medical Center in 1999. Between the years 2000 and 2002 he has obtained two clinical fellowships in Thoracic Imaging and Cardiac MRI and CT in Toronto General Hospital, Canada. In 2003, Prof. Konen established the first cardiac MRI service in Israel which developed since then to be the largest multidisciplinary cardiac Imaging service in Israel. In 2005, he accomplished his 2nd degree in Health Administration, Recanati Business Faculty, Tel Aviv University. Since 2005, he serves as the Head of

Diagnostic Imaging Department in Sheba Medical Center. Between 2012 and 2016 he served as the head of Imaging department in Sacker School of Medicine at Tel Aviv University. In the last five years Prof. Konen is involved in several projects regarding computerized medical image analysis: he established the first CILAB (Computational Imaging Laboratory) in Israel which is leading several original research projects as well as collaborating with other universities and the industry.

Michal Marianne Amitai received the M.D. degree in medicine from Tel Aviv University in 1983. Staff radiologist at the Chaim Sheba Medical Center since 1994. Head of CT service and Abdominal Imaging since 2011. Active in patient care, teaching and research. Senior Lecturer at Tel Aviv University. She is author or coauthor of more than 60 original articles and over 70 presentations in national and international conference. Her clinical strengths are in various fields of Radiology particularly Gastrointestinal and liver imaging. The two main research domain in recent years are (1) Automated image analysis with emphasize on automatic liver lesion detection and characterization in collaboration with the bio medical engineering department in Tel Aviv university. (2) Imaging of small bowel in Crohn's disease patient and correlation between video capsule endoscopy and MR enterography in the GIT imaging area. Dr. Amitai is also involved in the Israeli Board Examinations in Diagnostic Radiology being member of the Board Examination Committee and serves as member of the scientific and organizing committee of the annual International Meeting of the Israeli Radiological Association. She is also a very active member in the scientific committee of the Franco-Israelienne Annual Meeting dImagerie Mdicale (AFIIM) and also the annual French-Israeli Course in Radiology.

Hayit Greenspan is a Tenured Professor at the Biomedical Engineering Department in the Faculty of Engineering, Tel-Aviv University. She is affiliated with the International Computer Science Institute (ICSI) at Berkeley, CA. From 2008 until 2010, Dr. Greenspan was a visiting Professor at Stanford University, Department of Radiology, Faculty of Medicine. Dr. Greenspan has been conducting research in medical image analysis for the past 20 years, with a special focus on image modeling and analysis, deep learning, and content-based image retrieval. Dr. Greenspan received the B.S. and M.S. degrees in Electrical Engineering (EE) from the Technion - Israel Institute of Technology, and the Ph.D. degree in EE from CALTECH California Institute of Technology. She was a Postdoc with the Computer Science Division at U.C. Berkeley following which she joined Tel-Aviv University and founded the Biomedical Image Computing lab, which she heads. Among her current research projects: MRI resolution augmentation, Brain MRI research (structural and DTI), CT and X-ray image analysis and medical imagery indexing and retrieval. Dr. Greenspan has over 100 publications in leading international journals and conference proceedings. She has received several awards and is a coauthor on several patents. She has received grants from several Israeli and U.S. agencies. Currently her Lab is funded for Deep Learning in Medical Imaging by the INTEL Collaborative Research Institute for Computational Intelligence (ICRI-CI). Dr. Greenspan is member of several journal and conference program committees, including SPIE medical imaging, IEEE_ISBI and MICCAI. She is an Associate Editor for the IEEE Transactions on Medical Imaging (TMI) journal.

Please cite this article as: A. Ben-Cohen et al., Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.10.001