

# Towards Detecting High-Uptake Lesions from Lung CT Scans Using Deep Learning

Krzysztof Pawełczyk<sup>1,2</sup>, Michał Kawulok<sup>1,2</sup>, Jakub Nalepa<sup>1,2</sup>, Michael P. Hayball<sup>1,3</sup>, Sarah J. McQuaid<sup>4</sup>, Vineet Prakash<sup>5</sup>, and Balaji Ganeshan<sup>3,6</sup>

<sup>1</sup> Future Processing, Poland

<sup>2</sup> Silesian University of Technology, Poland

<sup>3</sup> Feedback PLC, UK

<sup>4</sup> Royal Surrey County Hospital, UK

<sup>5</sup> Ashford and St. Peter's Hospitals, UK

<sup>6</sup> University College London, UK

{krzysztof.pawelczyk, michal.kawulok, jakub.nalepa}@polsl.pl

**Abstract.** Automatic detection of lung lesions from computed tomography (CT) and positron emission tomography (PET) is an important task in lung cancer diagnosis. While CT scans make it possible to retrieve structural information, PET images reveal the functional aspects of the tissue, hence combined PET/CT imagery allows for detecting metabolically active lesions. In this paper, we explore how to exploit deep convolutional neural networks to identify the active tumour tissue exclusively from CT scans, which, to the best of our knowledge, has not been attempted yet. Our experimental results are very encouraging and they clearly indicate the possibility of detecting lesions with high glucose uptake, which could increase the utility of CT in lung cancer diagnosis.

**Keywords:** PET/CT imaging, lesion detection, deep neural networks

## 1 Introduction

Cancer is one of the main causes of death worldwide, with 1.69 out of 8.8 million deaths caused by lung cancer in 2015. Therefore, improving lung cancer diagnostics from medical images is crucial, either as a part of the screening process or at a later stage to assess the effectiveness of treatment. An important, yet challenging, task is to differentiate between benign and malignant lesions. Some of the useful features concerned with the structure, shape and boundary smoothness can be observed relying on *structural imaging* (such as computed tomography—CT or magnetic resonance imaging—MRI). However, in many cases the physiological activity of the tissue must be captured to make this differentiation, which can be achieved with *functional imaging*, such as positron emission tomography (PET), functional MRI or dynamic contrast-enhanced imaging.

PET imaging allows for measuring the glucose uptake, which indicates the metabolism of the tissue to identify abnormally active lesions. In lung cancer diagnostics, measuring the lesion's activity plays a key role in differentiating

between benign and malignant tumours or nodules. This provides functional information, however as PET images are of poor spatial resolution and they do not reveal much of the anatomical details, they are usually complemented with the co-registered CT scans. This is necessary, as the high-uptake regions (*hot spots* in PET) include many false positives (FPs) which can be easily verified based on the anatomy and structure (e.g., the heart is usually a hot spot).

### 1.1 Contribution

Taking into account that different type of information is acquired with CT and PET imagery, these modalities are often fused to improve the diagnosis. In the research reported here, we explore the possibility of detecting the high-uptake lesions exclusively from CT scans. Identifying active tumour tissue from CT would be very useful as this is an important indicator of disease severity and response. To the best of our knowledge, there is no reported work on this problem and PET images are considered indispensable here.

Our contribution consists in proposing a deep convolutional neural network (CNN) to detect and segment the high-uptake lesions from CT scans. Deep neural networks (DNNs) [14] have been already successfully applied to solve a number of computer vision problems, including medical imaging challenges [15], and in many cases they reach beyond human performance. We validated the outcome obtained from CT scans using our CNN against the annotations performed by professional radiologists, asked to locate the high-uptake regions based on CT and PET data. Although the detection scores are below the scores obtained using the combined CT and PET modality, they are highly encouraging and they suggest that DNNs are capable of identifying high-uptake lesions from structural images. Finally, we compared the detection results of the CNN with those reported by our recent PET/CT lung lesion detection algorithm (LUNGCTX) [17], which benefits from the information extracted from both modalities.

### 1.2 Paper Structure

This paper is structured as follows. Section 2 reviews the literature. The proposed CNN is described in Section 3 and the obtained experimental results are reported and discussed in Section 4. The paper is concluded in Section 5.

## 2 Related Literature

Detecting lung lesions (including nodules and tumours) is a deeply investigated problem of medical image analysis and there are numerous methods which operate from different image modalities. Here, we briefly outline the state of the art on detecting and segmenting lung lesions relying on CT and PET/CT images.

The two general tasks concerned with analysis of lung CT scans are: (i) detection of lung nodules aimed at early diagnosis of lung cancer and (ii) segmentation of the lesion region [11], helpful in differentiating between benign and malignant

lesions, computer-aided surgery or planning radiation treatment. Clinical aspects of detecting lung nodules from CT scans are thoroughly discussed in a recently published survey [1]. Sensitivity of state-of-the-art methods vary significantly, spanning between 80% [22] and 98% [5] at several FPs per whole CT scan. This allows for improving the performance of inexperienced radiologists, and in some cases increases the detection sensitivity when used by experienced radiologists.

Most of the existing lung nodule detectors determine a set of candidates, i.e., the regions of dense tissue inside the lungs, which are further verified to filter out the FPs. In [5], the nodule candidates are extracted following a number of simple rules and preliminarily verified in each 2D slice. Subsequently, the candidates are combined in 3D to extract textural features, and classified using support vector machines (SVMs). A similar approach, employing shape descriptors, was proposed in [20]—all of the nodules whose size is at least 10 mm were reported to be correctly detected at 4 FPs per scan. Recently, a deep CNN with 5 convolutional and 3 max pooling layers was applied to detect lung nodules [6]. The reported sensitivity is 78.9% at 20 FPs per scan without using any FP reduction.

The differential diagnosis of malignant from benign lesions is difficult from CT scans, if a tumour is built of soft tissue without calcifications, as the metabolic information is not known to be manifested in CT scans. PET imaging allows for measuring the concentration of biologically active molecules (usually fluorodeoxyglucose, FDG) marked with positron-emitting isotopes. Hence, the high-uptake regions can be identified to indicate the tissue of high metabolism, which is a well-known marker in cancer diagnosis. The PET intensities are converted into standardised uptake values [3] and verified taking into account the anatomy, extracted from CT scans [7, 17] or MRI [19]. The malignant lesions usually are seen as hot spots in PET due to increased metabolism of a tumour, but the hot spots do also appear in the healthy tissue (e.g., in the heart). In [7], the entire-body CT scan is divided into several sections using hidden Markov model, which subsequently makes it possible to classify each hot spot as normal or abnormal.

Not only are the CT scans used to verify the hot spots based on human body atlas, but they also allow for increasing the precision of delineating the lesions. The hot spots extracted from PET are treated as seeds for image segmentation performed with numerous techniques, including graph cuts [2], Markov random fields [8] or random walks [10]. The information extracted from PET may also be used during segmentation [25]. In [4], local maxima and saddle points are detected in a PET image to create a spatial-topological distance map, from which the tumours are segmented. In [23], the nodules are detected independently in both modalities—as hot spots in PET and using active contour filters in CT scans, and then a CNN with 3 convolutional layers is applied to extract the features of each candidate, which are classified with an SVM. Adding the CNN-based verification allowed for reducing the number of FPs from 72.8 to 4.9 per case, while the sensitivity dropped from 97.2% to 90.1%.

Deep CNNs have been successfully used for detecting and segmenting lung lesions both from CT and PET/CT modalities [6, 23], but we have not encountered any reported attempts to bridge the gap between the results obtained from

CT scans alone and from the combined PET/CT modalities. Since the deep networks allow for reaching beyond human performance in certain computer vision tasks, they may be helpful in detecting potentially high FDG uptake lesions from CT scans, improving the utility of CT in lung cancer diagnosis.

### 3 Detecting High-Uptake Lesions Using a CNN

Our algorithm for detecting active lesions is outlined in Fig. 1. At first, the 16-bit pixel values are normalised to the range of  $(0, 1)$  and the CT scan is split into 3D patches of size  $5 \times 75 \times 75$  (2D patches of size  $75 \times 75$  pixels are retrieved from 5 subsequent slices, as shown in Fig. 2). Such patches are used to train our deep CNN and afterwards each patch is classified by the trained network as *lesion* or *background*. For each patch, the CNN returns two responses ( $r_l$  and  $r_b$ ) that express the similarity of the patch central pixel to the lesion and background classes, respectively. From these responses, the lesion similarity map is assembled to determine the lesion candidates, which are subject to two verification steps, based on (i) the maximal similarity within a blob and (ii) the blob’s area.

**Network Architecture.** The proposed network, whose architecture is presented in Fig. 3, is composed of two convolutional layers followed by three classical fully connected layers. According to [21], for small input images the pooling may be skipped to increase the performance, therefore there are no pooling layers in our CNN. Output of every hidden layer is adjusted by a rectified linear unit (ReLU) [16] with an activation function  $f(x) = \max(0, x)$ . The main task of the first convolutional layer is to extract the low-level features from adjacent slices using a bank of 16 filters of the size  $5 \times 9 \times 9$  pixels, applied with a stride of 2 pixels (the size of each patch is reduced from  $5 \times 75 \times 75$  to  $34 \times 34$ ). 3D

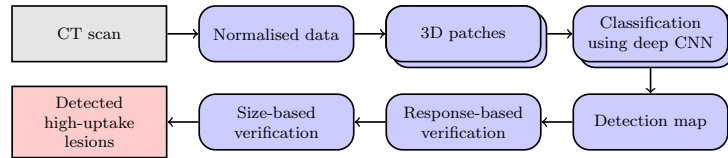


Fig. 1. Flowchart of the proposed DNN-based high-uptake lesions detection.

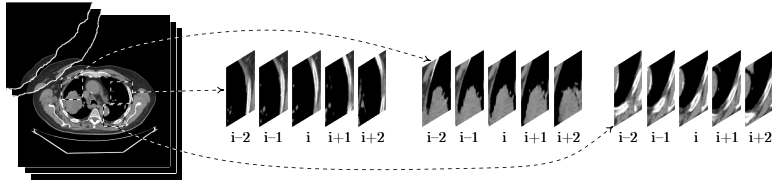
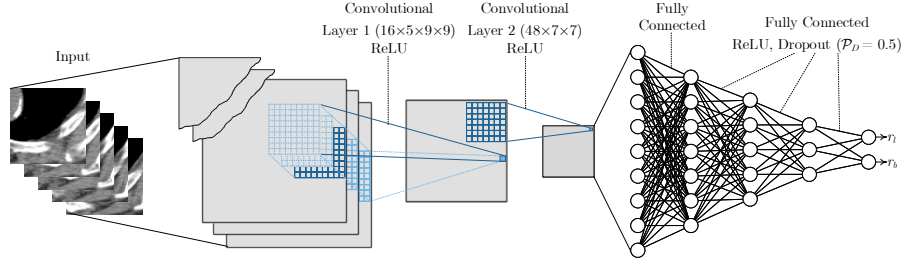


Fig. 2. Examples of 3D patches extracted for the  $i$ -th slice.



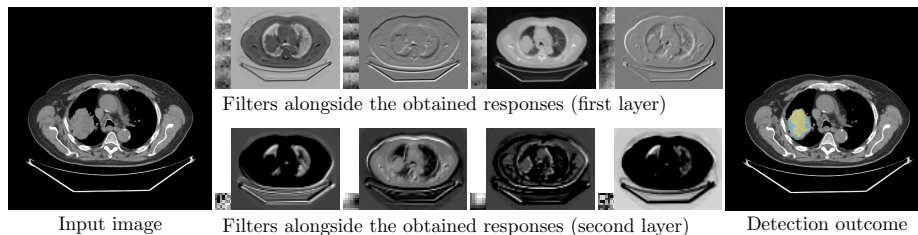
**Fig. 3.** Architecture of the proposed deep CNN.

patches are transformed into 2D ones of reduced dimensionality here. The second convolutional layer with the bank of 48 kernels of size  $7 \times 7$  with a stride of 3 extracts higher-order features. The fully connected layers are a classical neural network with dropout (its probability is  $\mathcal{P}_D$ ) which classifies the input vector.

**Handling Extremely Imbalanced Data Sets.** Training classification engines from extremely imbalanced data sets is a vital research topic, since the skewed distributions of examples may easily bias the classifiers [13]. This problem is commonly addressed in deep learning with data augmentation and undersampling—these procedures are usually executed before the training, to learn the CNN using a new, potentially “balanced” set. Hence the training process is often unaware of the underlying data characteristics and cannot adapt to retrieve a better-performing model. This was thoroughly discussed in a very recent work [24].

In our approach, we dynamically balance the data batches (each contains  $N$  samples) during the CNN training. After every  $\mathcal{I}$  epochs of the CNN training, the classification accuracy  $\eta$  is quantified for each class separately (the validation set is balanced, and encompasses all minority-class examples, along with under-sampled examples from the other class). Consequently, the lower  $\eta$  retrieved for a given class, the higher probability of including its samples in the batch.

**Illustrative Examples of CNN Filters.** Several examples of the learned filters (corresponding to the architecture presented in Fig. 3), along with the filtered images, are presented in Fig. 4 (the first layer responses are visualised before applying ReLU, while those for the second layer—after ReLU). For the first layer, each filter is composed of five 2D filters convolved with subsequent CT slices. It can be seen that these filters do not resemble the wavelets often reported in the literature, and they are rather “noisy”—possibly, some textural features are extracted in this way. Interestingly, the filters at the second layer are smoother and they are focused on extracting higher-order features, as expected. The extracted features in the presented example allow for correct segmentation of an active lesion (yellow region in the detection outcome indicates true positives, while the blue—false negatives). We have observed that nearly half of



**Fig. 4.** Examples of filtered images obtained at the first and second convolutional layer.

the learned filters are uniformly gray when visualised, hence they average the signal—presence of such “dead” filters may mean that there are too many of them in the layer or that they should be of smaller dimensions [27]. Addressing that problem may improve our approach in the future.

**Analysis of the Network Response.** From the pixel-wise responses  $r_l$  and  $r_b$ , we compute the lesion similarity ( $\mathcal{R} = r_l - r_b$ ) to create the lesion detection map. The pixels with non-negative similarities ( $\mathcal{R} > 0$ ) are grouped spatially and each consistent region is considered a lesion candidate. These candidates are verified using two thresholds  $\mathcal{T}_R$  and  $\mathcal{T}_S$ , imposed on the blob’s maximum similarity ( $\mathcal{R}_{max}$ ) and its area in pixels ( $\mathcal{S}$ ), respectively. If  $\mathcal{R}_{max} > \mathcal{T}_R$  and  $\mathcal{S} > \mathcal{T}_S$ , then the blob is labeled as the detected lesion.

## 4 Experimental Results

### 4.1 Experimental Setup

We validated our algorithm using two data sets, namely (i) our set with 90 CT scans of different patients (this includes the LUNG CX subset—44 scans used in [17]) and (ii) the LOLA set<sup>7</sup> with 55 CT scans without active lesions. For every study in our set (with active and non-active lesions), a single slice presenting the largest section of an active lesion was selected and manually segmented by an experienced radiologist based on both PET and CT (our algorithm *does not* exploit PET images). This set is extremely imbalanced—there are  $7.3 \cdot 10^4$  pixels of active lesions, and  $2.4 \cdot 10^7$  pixels of other tissues and background.

The algorithms were implemented in C++ with the Caffe framework [9] and validated on a computer with an Intel Xeon E5-2698 v3 processor (40M Cache, 2.30 GHz) with 128GB RAM and NVIDIA Tesla K80 GPU 24GB DDR5. The CNN internals were tuned to  $\mathcal{P}_D = 0.5$ ,  $N = 256$  and  $\mathcal{I} = 500$ , and we use the ADAM optimizer [12]. As there are no reported attempts to detect high-uptake lesions from CT, we compare our method with LUNG CX [17], exploiting the combined PET/CT modality. The reported results were obtained with 10-fold

<sup>7</sup> LOLA set is available at <https://lola11.grand-challenge.org>.

cross-validation. The LOLA set was processed  $10\times$  for CNN trained within each fold using our data. Processing a single slice consumes 135 s on average.

## 4.2 Analysis and Discussion

**Quantitative Analysis.** Table 1 presents the obtained detection scores. For our data set, we report precision and recall for the training and test sets (including the LUNG CX subset), averaged over 10 folds (entire slices are segmented). As the LOLA scans do not include active lesions, we only report the FP rate (i.e., the percentage of images with FP lesions). We show the scores obtained (i) without any verification ( $\mathcal{R} > 0$ ), (ii) after response-based verification, (iii) with size-based verification, and (iv) after full verification. Clearly, the verification is critical, as it drastically decreases the FPs for LOLA (from 90.14% to 6.6%) and significantly improves the precision. The F-score differences are statistically important for all data sets at  $p < .01$  (two-tailed Wilcoxon test). We also report the recall for four ranges of lesion size. Although the verification decreases the recall, mainly small lesions are affected—while this is obvious for size-based verification, the smaller active lesions also render lower  $\mathcal{R}$ , hence their vulnerability to the response-based cutoff.

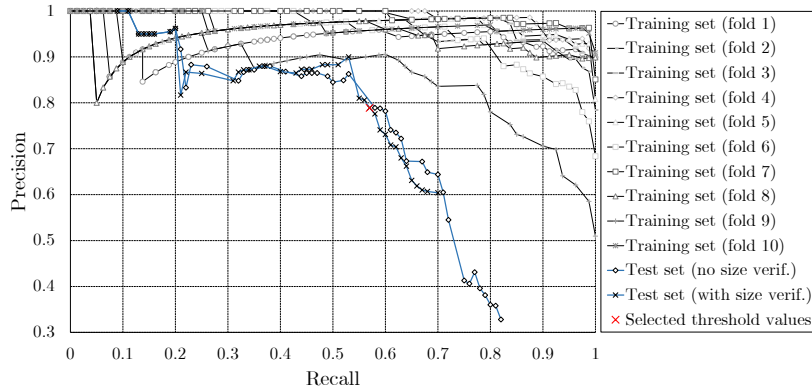
In Figs. 5 and 6, we show the precision-recall curves obtained by varying the thresholds  $\mathcal{T}_R$  and  $\mathcal{T}_S$ . The curves for the training set differ much across the folds, so increasing the amount of training data could be beneficial. For the test set, we apply either single or both thresholds (for each case only one threshold is being changed, while the other remains fixed), hence two curves for each case. The fixed threshold values were found independently for each fold, so as to obtain equal precision and recall for the training set, and these values were also applied to obtain the scores reported earlier in Table 1. While applying  $\mathcal{T}_R$  after  $\mathcal{T}_S$  improves the results (Fig. 6), the latter does not render any improvement, if the former is applied (Fig. 5). Overall, we use  $\mathcal{T}_R$  and  $\mathcal{T}_S$ , as we observed (from LOLA) that  $\mathcal{T}_S$  is quite effective in reducing FPs for lungs without active lesions.

**Table 1.** Precision and recall obtained for our data set and FP rate for the LOLA set.

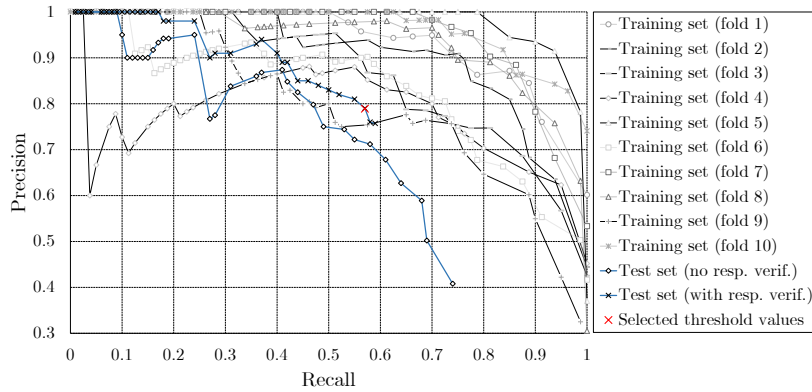
Verif.	Precision	Recall	Recall for lesions of size (in mm)				W/out active FP rate (%)
			< 15	(15, 25)	(25, 40)	> 40	
Training set:							
None	.278 ± .102	1 ± .000	1 ± .000	1 ± .000	1 ± .000	1 ± .000	—
Resp.	.905 ± .052	.911 ± .050	.747 ± .115	.956 ± .053	.971 ± .060	1 ± .000	—
Size	.814 ± .060	.840 ± .074	.442 ± .242	.984 ± .034	1 ± .000	1 ± .000	—
Full	.925 ± .038	.808 ± .078	.379 ± .213	.951 ± .065	.971 ± .060	1 ± .000	—
Test set (full):							
None	.187 ± .072	.870 ± .064	.574 ± .259	.917 ± .171	1 ± .000	.981 ± .052	90.14 ± 8.14
Resp.	.741 ± .192	.610 ± .170	.269 ± .232	.758 ± .342	.622 ± .171	.867 ± .180	20.42 ± 9.68
Size	.589 ± .130	.680 ± .166	.315 ± .196	.758 ± .259	.811 ± .241	.944 ± .111	26.04 ± 9.11
Full	.789 ± .179	.570 ± .179	.176 ± .173	.708 ± .340	.622 ± .171	.852 ± .183	6.60 ± 3.52
Test set (LUNG CX subset):							
None	.261 ± .099	1 ± .000	—	1 ± .000	1 ± .000	1 ± .000	90.14 ± 8.14
Resp.	.852 ± .135	.677 ± .272	—	.700 ± .400	.593 ± .284	.852 ± .319	20.42 ± 9.68
Size	.779 ± .160	.872 ± .137	—	.700 ± .400	.824 ± .210	1 ± .000	26.04 ± 9.11
Full	.935 ± .134	.677 ± .272	—	.700 ± .400	.593 ± .284	.852 ± .319	6.60 ± 3.52

**Qualitative Analysis.** Fig. 7 shows several examples of correct (Fig. 7a–e) and incorrect (Fig. 7f–i) detection. A very interesting example is presented in Fig. 7c—although the lesion is not very dense, our CNN identified it correctly and managed to differentiate it from another non-active lesion, which is present in the image. In Fig. 7f, the lesion was not detected and there has been one FP region found—naturally, we consider such cases as detection errors. Fig. 7g and (h) show the outcome before and after the verification—several FPs were eliminated from (g), but the correctly detected lesion in (g) was also rejected.

**Comparison with LUNG CX.** We compared the proposed CNN classifier with our active lesion detection algorithm (LUNG CX) which operates on both PET and CT modalities [17]. In LUNG CX, the co-registered CT and PET series are identified at first, and the lungs (base and apex) are located from CT in the pixel-intensity histogram analysis. Then, for each lung-containing slice, we



**Fig. 5.** Precision-recall curves obtained by varying the response threshold  $\mathcal{T}_R$ .



**Fig. 6.** Precision-recall curves obtained by varying the size threshold  $\mathcal{T}_S$ .

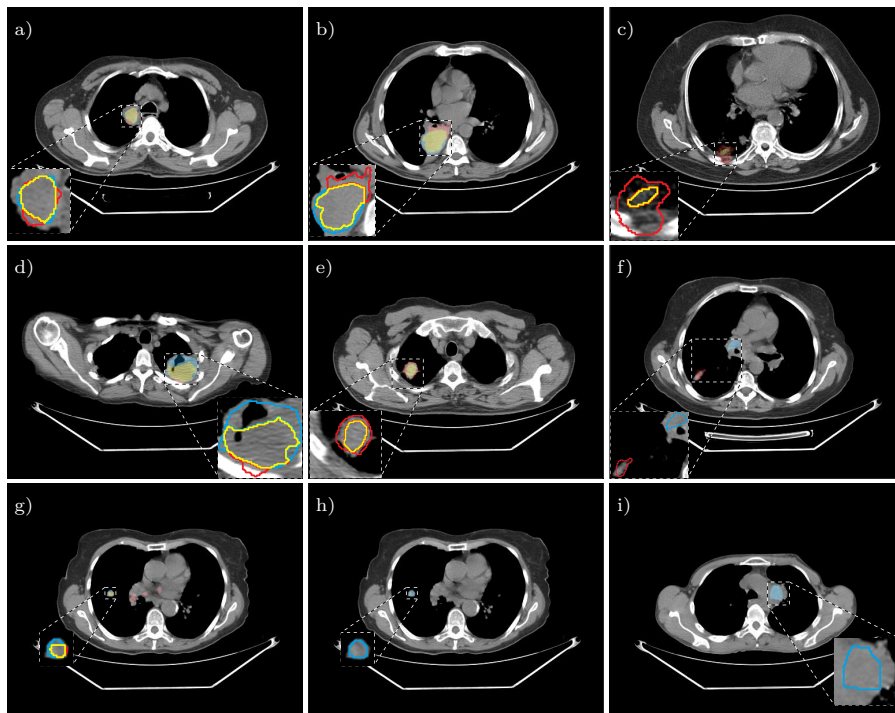


identify the lung tissue using thresholding, alongside its convex hull, as tumours may be associated with the lung wall or mediastinum. The active lesions (located only within the convex hulls of lungs) are extracted from the PET images [26].

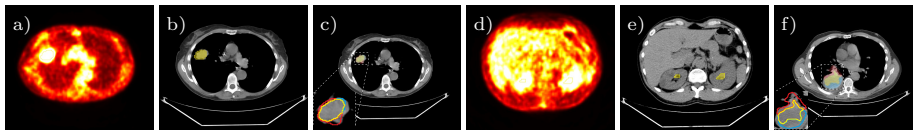
Although the LUNGEX algorithm successfully identified active lesions in 41 (out of 44 patients in the LUNGEX subset) cases (93%), the example visualisations gathered in Fig. 8 show that the most avid PET regions are very patient-specific. In Fig. 8, we render the example of the problematic uptake in which the most active regions were found in the kidneys (Fig. 8d–e). Here, our CNN detected the active lesion correctly (Fig. 8f) and did not report any lesions for these high-uptake kidneys. The numerical results (Table 1) reveal that our CNN can produce results comparable with LUNGEX, and the analysis of the network response (see the Size variant) greatly improves the CNN recall measures.

## 5 Conclusions and Outlook

In this paper, we reported our attempt to employ deep learning for detecting high-uptake lesions from CT images, which is a challenging task, as it requires extraction of functional information from structural imaging. The obtained detec-



**Fig. 7.** Examples of correct (a–e) and incorrect (f–i) active lesion detection (yellow: true positives, red: false positives, blue: false negatives).



**Fig. 8.** Examples of active lesion detection from PET/CT with LUNG CX (a, b, d, e) and from CT using the proposed algorithm (c, f).

tion scores, though worse than those retrieved from the combined PET/CT [17], are encouraging and they indicate that our algorithm correctly differentiates between active and non-active lesions. It is clear from the experimental results that methods utilising deep CNNs can increase the CT diagnostic capacity.

Our ongoing research is aimed at improving the visualisation aspects to better understand which features are learned by the CNN and which image regions activate them, alongside comparing our method with other state-of-the-art techniques on larger data sets. Furthermore, we intend to focus on detecting small active lesions, being an important clinical goal, especially from low-dose CT. We are on the way to employ the existing methods for CT-based lesion detection (also exploiting the anatomical information) to pre-process the scans and narrow down the search in the pulmonary region. Also, we work on applying incrementally increased CNN architectures in our framework [18]. Overall, while the proposed method could be improved on many ways, it is an important step towards retrieving information on lesion activity from CT.

**Acknowledgments.** This work was supported by the National Centre for Research and Development under the grant: POIR.01.02.00-00-0030/15.

## References

1. Al Mohammad, B., Brennan, P., Mello-Thoms, C.: A review of lung cancer screening and the role of computer-aided detection. *Clinical Radiology* (2017)
2. Ballangan, C., Wang, X., Fulham, M., Eberl, S., Feng, D.D.: Lung tumor segmentation in PET images using graph cuts. *Computer Methods and Programs in Biomedicine* 109(3), 260–268 (2013)
3. Boellaard, R.: Mutatis mutandis: harmonize the standard! *Journal of Nuclear Medicine* 53(1), 1–3 (2012)
4. Cui, H., Wang, X., Lin, W., Zhou, J., Eberl, S., Feng, D., Fulham, M.: Primary lung tumor segmentation from PET-CT volumes with spatial-topological constraint. *Int. Journal of Computer Assisted Radiology and Surgery* 11(1), 19–29 (2016)
5. Demir, Ö., Çamurcu, A.Y.: Computer-aided detection of lung nodules using outer surface features. *Bio-Medical Materials and Engineering* 26(s1), S1213–S1222 (2015)
6. Golan, R., Jacob, C., Denzinger, J.: Lung nodule detection in ct images using deep convolutional neural networks. In: *Proc. IJCNN*. pp. 243–250 (2016)
7. Guan, H., Kubota, T., Huang, X., Zhou, X.S., Turk, M.: Automatic hot spot detection and segmentation in whole body FDG-PET images. In: *Proc. IEEE ICIP*. pp. 85–88 (2006)

8. Guo, Y., Feng, Y., Sun, J., Zhang, N., Lin, W., Sa, Y., Wang, P.: Automatic lung tumor segmentation on PET/CT images using fuzzy Markov random field model. *Computational and Mathematical Methods in Medicine* 2014 (2014)
9. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093* (2014)
10. Ju, W., Xiang, D., Zhang, B., Wang, L., Kopriva, I., Chen, X.: Random walk and graph cut for co-segmentation of lung tumor on pet-ct images. *IEEE Trans. on Image Processing* 24(12), 5854–5867 (2015)
11. Keshani, M., Azimifar, Z., Tajeripour, F., Boostani, R.: Lung nodule segmentation and recognition using SVM classifier and active contour modeling: A complete intelligent system. *Computers in Biology and Medicine* 43(4), 287–300 (2013)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR abs/1412.6980* (2014)
13. Krawczyk, B.: Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence* 5(4), 221–232 (2016)
14. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* 521(7553), 436–444 (2015)
15. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. on Medical Imaging* 35(11), 2369–2380 (2016)
16. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. *Proc. ICML* pp. 807–814 (2010)
17. Nalepa, J., Szymanek, J., McQuaid, S., Endozo, R., Prakash, V., Ganeshan, B., Menys, A., Hayball, M., et al.: PET/CT in lung cancer: An automated imaging tool for decision support. In: *Proc. RSNA*. pp. 1–2 (2016)
18. Ribalta Lorenzo, P., Nalepa, J., Kawulok, M., Sanchez Ramos, L., Ranilla Pastor, J.: Particle swarm optimization for hyper-parameter selection in deep neural networks. In: *Proc. GECCO*. pp. 1–8. ACM, USA (2017)
19. Rundo, L., Stefano, A., Militello, C., et al.: A fully automatic approach for multimodal pet and mr image segmentation in gamma knife treatment planning. *Computer Methods and Programs in Biomedicine* (2017)
20. Setio, A.A., Jacobs, C., Gelderblom, J., Ginneken, B.: Automatic detection of large pulmonary solid nodules in thoracic CT. *Med. Phys.* 42(10), 5642–5653 (2015)
21. Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: The all convolutional net. *ECCV 2014 arXiv preprint* (2014)
22. Teramoto, A., Fujita, H.: Fast lung nodule detection in chest CT images using cylindrical nodule-enhancement filter. *Int. J. of Comp. Assisted Radiol. and Surg.* 8(2), 193–205 (2013)
23. Teramoto, A., Fujita, H., Yamamuro, O., Tamaki, T.: Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Medical Physics* 43(6), 2821–2827 (2016)
24. Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., Kennedy, P.J.: Training deep neural networks on imbalanced data sets. In: *Proc. IJCNN*. pp. 4368–4374 (2016)
25. Wang, X., Ballangan, C., Cui, H., Fulham, M., Eberl, S., Yin, Y., Feng, D.: Lung tumor delineation based on novel tumor-background likelihood models in PET-CT images. *IEEE Trans. on Nuclear Science* 61(1), 218–224 (2014)
26. Win, T., Miles, K.A., Janes, S.M., et al.: Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung cancer. *Clinical Cancer Research* 19(13), 3591–3599 (2013)
27. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: *European conference on computer vision*. pp. 818–833. Springer (2014)