

Biomedical Physics & Engineering Express



PAPER

Liver lesion localisation and classification with convolutional neural networks: a comparison between conventional and spectral computed tomography

OPEN ACCESS

RECEIVED
22 October 2019

REVISED
26 December 2019





ACCEPTED FOR PUBLICATION
15 January 2020

PUBLISHED
31 January 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Nadav Shapira¹ , Julia Fokuhl^{2,3}, Manuel Schultheiß², Stefanie Beck², Felix K Kopp² , Daniela Pfeiffer², Julia Dangelmaier², Gregor Pahn⁴ , Andreas P Sauter², Bernhard Renger², Alexander A Fingerle² , Ernst J Rummeny², Shadi Albarquni^{3,5}, Nassir Navab³ and Peter B Noël^{1,2}

¹ Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, United States of America

² Department of Diagnostic and Interventional Radiology, Technical University of Munich, Munich, Germany

³ Computer Aided Medical Procedures (CAMP), Technical University of Munich, Garching, Germany

⁴ Clinical Science CT, Philips Germany GmbH, Hamburg, Germany

⁵ Department of Information Technology and Electrical Engineering, ETH, Zurich, Switzerland

E-mail: nadav.tech@gmail.com

Keywords: CNN, liver lesion, computed tomography, dual-energy CT, AI, hepatic imaging, spectral CT

Supplementary material for this article is available [online](#)

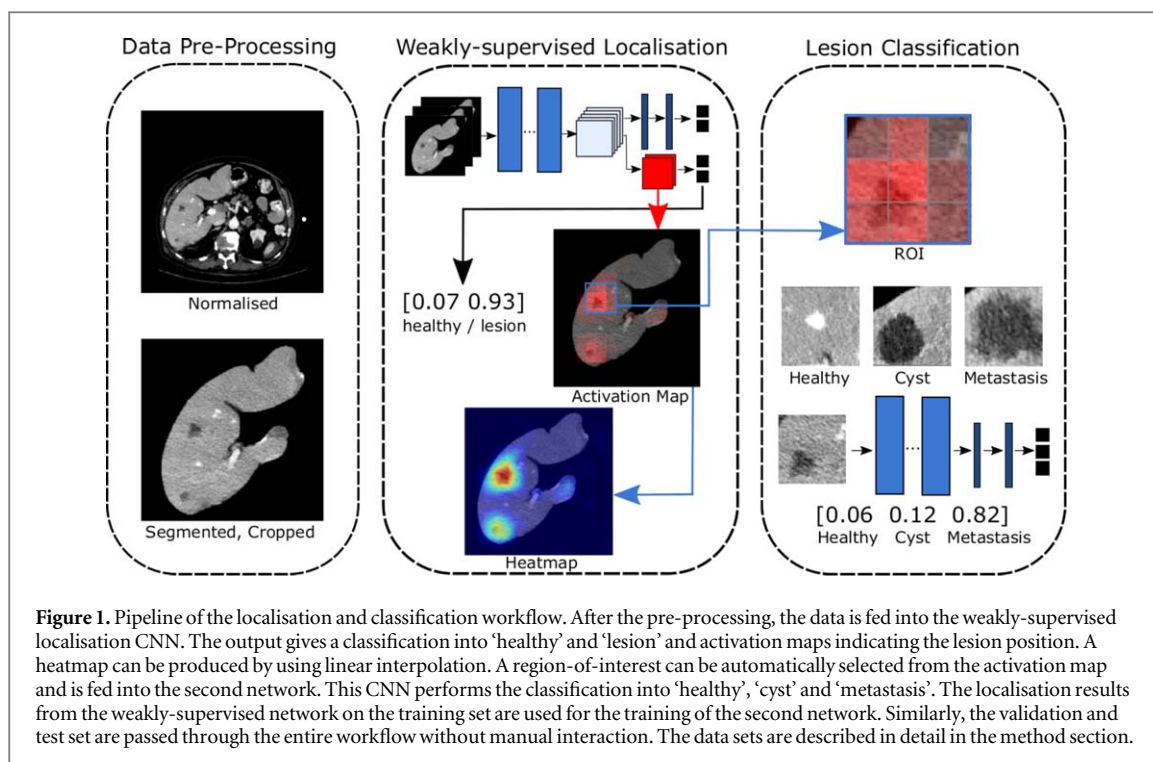
Abstract

Purpose: To evaluate the benefit of the additional available information present in spectral CT datasets, as compared to conventional CT datasets, when utilizing convolutional neural networks for fully automatic localisation and classification of liver lesions in CT images. **Materials and Methods:** Conventional and spectral CT images (iodine maps, virtual monochromatic images (VMI)) were obtained from a spectral dual-layer CT system. Patient diagnosis were known from the clinical reports and classified into healthy, cyst and hypodense metastasis. In order to compare the value of spectral versus conventional datasets when being passed as input to machine learning algorithms, we implemented a weakly-supervised convolutional neural network (CNN) that learns liver lesion localisation without pixel-level ground truth annotations. Regions-of-interest are selected automatically based on the localisation results and are used to train a second CNN for liver lesion classification (healthy, cyst, hypodense metastasis). The accuracy of lesion localisation was evaluated using the Euclidian distances between the ground truth centres of mass and the predicted centres of mass. Lesion classification was evaluated by precision, recall, accuracy and F1-Score. **Results:** Lesion localisation showed the best results for spectral information with distances of 8.22 ± 10.72 mm, 8.78 ± 15.21 mm and 8.29 ± 12.97 mm for iodine maps, 40 keV and 70 keV VMIs, respectively. With conventional data distances of 10.58 ± 17.65 mm were measured. For lesion classification, the 40 keV VMIs achieved the highest overall accuracy of 0.899 compared to 0.854 for conventional data. **Conclusion:** An enhanced localisation and classification is reported for spectral CT data, which demonstrates that combining machine-learning technology with spectral CT information may in the future improve the clinical workflow as well as the diagnostic accuracy.

Introduction

The liver is a common site for malignant and benign lesions. Primary liver cancer causes the 2nd and 6th highest number of estimated cancer deaths for men and women worldwide [1]; additionally, the liver is one of the most frequent sites for metastatic spread. There is a rising number of incidentally discovered

lesions in the liver [2], due to the increasing use and sensitivity of diagnostic imaging modalities such as ultrasound, magnetic resonance imaging (MRI) or computed tomography (CT). Although a majority of incidentally detected lesions are benign, the lesion type is often unclear at the time of initial discovery [3]. Benign lesions in the liver include hepatic cysts, which are abnormal fluid filled spaces inside the liver. They



are relatively common with an estimated incidence of 2.5% in the normal population [4], and usually do not present any symptoms [5]. Malignant lesions include primary cancers such as hepatocellular carcinoma (HCC) as well as metastases. 40%–50% of patients with a primary tumour will develop hepatic metastases [4]. The distinction between benign and malignant liver lesions is crucial for an appropriate and individual treatment of each patient.

In 2006, spectral or dual-energy CT was first introduced into the clinical routine [6, 7]. During a spectral CT acquisition, two datasets are recorded via different x-ray spectra (e.g. 80 kVp & 140 kVp) [8, 9] or by energy resolving detectors [10, 11]. This technology can provide information on the chemical composition of different tissues and materials in the human body (e.g. kidney stones [12]) by measuring the energy and material dependent attenuation coefficient. One spectral CT feature which allows improved detection, depiction and characterization of liver lesions in contrast enhanced scans is virtual monoenergetic images (VMIs) [13, 14]. VMIs mimic a CT scan acquired with a monoenergetic source and can be clinically calculated for the energy range 40–200 keV [15]. If image noise is treated appropriately, low VMIs (<70 keV) offer an improved iodine contrast and lesion to parenchyma contrast since the energy setting is moving closer to the k-edge of iodine [16, 17]. In addition, spectral curves representing HU values as a function of VMI energy values in the portal venous phase can be used quantitatively to classify whether tumors are benign or malignant [13].

An early implementation of computer-aided diagnostic (CAD) systems to support the classification of

liver lesions can be found in Gletsos *et al* [18]. Here, a hand-crafted feature vector is used to discriminate between normal liver, hepatic cysts, hemangiomas and hepatocellular carcinomas. Starting in computer vision, convolutional neural networks (CNNs) were shown to perform well at advanced tasks such as object classification [19]. Classification of liver lesions can also be performed with CNNs [20]. However, the input to the network usually needs to be a regions-of-interest (ROIs) around the lesion manually selected by a radiologist. Additionally, deep neural networks require a vast number of training samples to optimise all parameters. In the medical domain however, there will often only be a limited amount of labelled samples to train with. Therefore, it is common to build new networks on top of pre-trained architectures such as AlexNet or Resnet-50 [19, 21]. Another difficulty in the medical domain is the acquisition of ground truth annotations, as they are usually provided by medical experts. If only image-level labels are available, weakly-supervised learning can still produce localisation [22, 23]. A common approach is to use the activation of a layer within the network to find the part of the input image that influences the network's decision the most [24, 25].

Automatic liver lesion classification can potentially benefit from spectral CT data [26]. The mean intensities of the lesions vary for different VMIs and especially at lower energies, the contrast between the different types of lesions is significantly increased [13, 14]. In the current work, the benefits of spectral CT in comparison to conventional CT for fully automatic localisation and classification of liver lesions with different weakly-supervised CNN models is

evaluated. The proposed CNN-based method consists of an initial stage for localisation of anomalies and a second stage for classification of liver lesions (see figure 1).

Methods

Data set and pre-processing

In this retrospective and IRB approved study, all CT images were selected from the Picture Archiving and Communication System (PACS). The data was acquired with a dual-layer spectral CT (IQon Spectral CT, Philips Healthcare, Eindhoven, The Netherlands) from 2016–2017. VMIs from 40–140 keV and iodine concentration maps were calculated. Conventional and spectral reconstructions of each dataset were transferred to a research environment for software development (IntelliSpace Discovery, Philips Healthcare, Eindhoven, The Netherlands).

All CT scans were acquired with contrast enhancement using standard clinical iodine contrast agent (Imeron 400, Bracco, Konstanz, Germany). The images were taken during the portal venous phase (delayed phase) 70 s after complete contrast agent injection with a slice thickness of 0.9 mm. The scans were performed with 120 kVp, 0.984 pitch, 0.33 s rotation time, 64×0.625 mm collimation and 3D dose modulation, resulting in an average CTDIvol value of 10.6 ± 3.5 mGy and an average DLP value of 353.0 ± 137.85 mGy. All images had a matrix size of 512×512 , the field of view was adjusted to patient size, resulting in different resolutions. For the test set, the in-plane pixel size ranged from 0.68 mm to 0.98 mm. For all images, a liver segmentation tool available in the IntelliSpace Discovery platform and based on the algorithm described in [27] was used to segment the livers automatically. The segmentation mask, which was obtained using the conventional images, was applied on all corresponding spectral images of the patient. The segmented liver images were pre-processed before being used in the neural network. All conventional and VMIs were clipped between -100 and 400 HU providing an ideal windowing for the liver [28]. The values were then normalised to $[0,1]$. The iodine maps were normalised to $[0,1]$ without clipping, unless strong artefacts appeared. Finally, the image was cropped to contain the whole liver with a small border and all slices were rescaled to 224×224 pixels in order to generate the correct input size for the pre-trained network (see figure 1).

Patient population

This study included data from 172 different patients, selected from the PACs system by a keyword search. 33 patients with one or multiple cysts, 57 with one or multiple hypodense metastases and 82 patients without liver lesions were included where all cases were verified to meet the selection criteria by an expert

radiologist. The same expert radiologist also re-confirmed the correctness of each lesion classification as it appears in the original clinical report. For patients with lesions, only slices containing a part of a lesion were used during the training. Although the initial image dataset did include a single case that contained multi-labelled images, this case was disregarded from final dataset used in the study. This yielded a total of 1187 slices with cysts, 5226 slices with metastases and 12236 slices with healthy liver tissue. The available patient data was split into a 60% training, 20% validation and 20% test set. The slices of an individual patient were always fully contained in one set. Due to the highly imbalanced number of slices per class, only parts of the whole dataset were used. For the cyst class, all available slices were used. For the metastasis and healthy class, slices were selected automatically always skipping a fixed number of slices. For the training of the networks, the classes were balanced to contain 50% healthy cases and 50% lesion cases. The slices with lesions were composed of 50% cysts and 50% metastases slices, resulting in 1439, 704 and 740 slices for healthy, cyst and metastasis cases in the training set, respectively.

The test set contained a total of 739 slices that were used for the final evaluation of the network after training. These images originated from 17 patients without lesion, 5 with cysts and 11 with metastases. The numbers of slices for the healthy, cyst and metastasis class were 255, 252 and 232, respectively. Slices with cysts had between 1 and 7 lesions per slice and slices with metastases had between 1 and 10 lesions per slice. The size of the cysts ranged between 3–70 mm, metastases measured between 3–108 mm. The lesions in the test set were segmented manually and verified by an experienced radiologist (6 years of experience) to serve as a ground truth for evaluating the accuracy of lesion localizations produced by the weakly-supervised CNN algorithm.

Weakly-supervised convolutional neural network

The input to the algorithm contains axial conventional or spectral CT slices of segmented livers (see figure 1, Data-Pre-processing). The algorithm is composed of three parts as follows. In the first part, two branches emerge from a pre-trained CNN where the first branch classifies the image into ‘healthy’ or ‘lesion’ and the outcome of the second branch are activation maps. The activation map for the lesion landmarks alone is further used. The second part of the algorithm acts as an automatic ROI selection tool using the activation maps of lesions as input. Both single instance learning (SIL) and multiple instance learning (MIL) methods were compared where in SIL a single ROI was selected based on the location of maximum value within the activation map and in MIL three ROIs were selected based on clipping the activation map at 70% of its maximal value [24] and analysing the resulting

Table 1. Classification Results.

Input Data Type	F1-Score		Precision		Recall		Accuracy Overall
	Healthy	Lesion	Healthy	Lesion	Healthy	Lesion	
40 keV	0.861	0.920	0.815	0.951	0.914	0.890	0.899
50 keV	0.833	0.904	0.789	0.934	0.882	0.876	0.878
60 keV	0.830	0.902	0.786	0.932	0.878	0.874	0.876
70 keV	0.805	0.887	0.758	0.920	0.859	0.855	0.857
100 keV	0.793	0.865	0.703	0.944	0.910	0.798	0.836
140 keV	0.802	0.872	0.715	0.947	0.914	0.808	0.844
Iodine Map	0.845	0.909	0.791	0.946	0.906	0.874	0.885
Conventional	0.812	0.881	0.730	0.948	0.914	0.822	0.854

Classification results for the test dataset containing 739 slices from 33 patients comparing different spectral and conventional data for the weakly-supervised CNN. For the definition of the different metrics and details on the test set see the method section.

connected components. In the final part of the algorithm, a CNN with a low number of trainable parameters was trained from scratch to yield three outputs which, after applying a softmax function, constitutes the class prediction. For MIL three CNN branches were trained in the final part of the algorithm, each receiving one of the three input ROIs.

CNNs were trained and tested separately for the conventional and each of the spectral results where the workflow for the testing dataset is the same as for the training and validation datasets (as shown in figure 1). All the results in the following sections are reported for test datasets. A more detailed overview of development and implantation of our algorithm for localization and classification can be found in the supplemental material is available online at stacks.iop.org/BPEX/6/015038/mmedia.

Metrics

For the evaluation of the classification, the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) were counted and used to calculate precision, recall, accuracy and F1-score.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

$$\text{F1 - Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

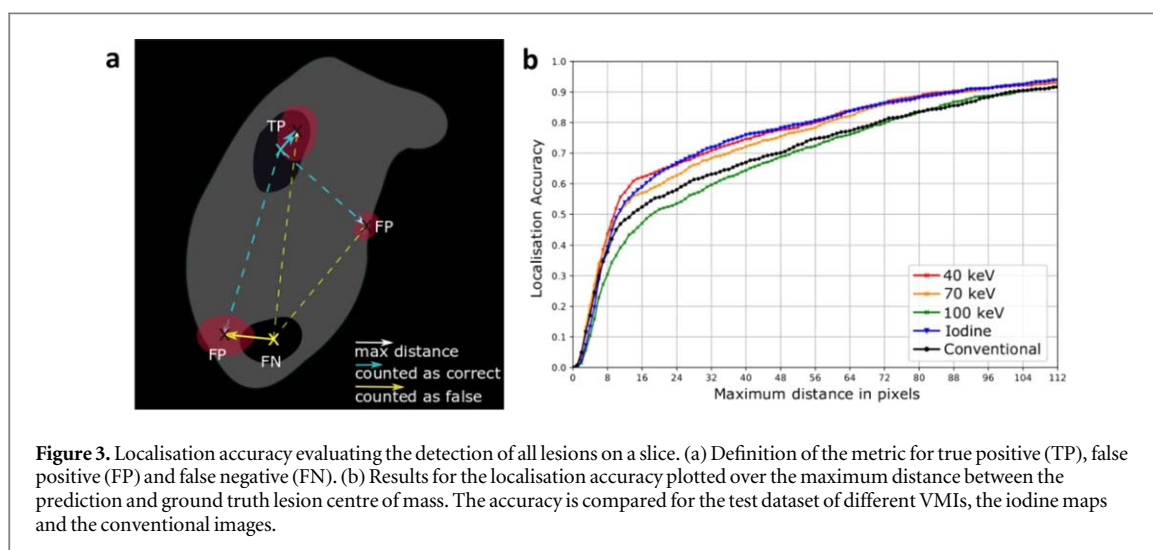
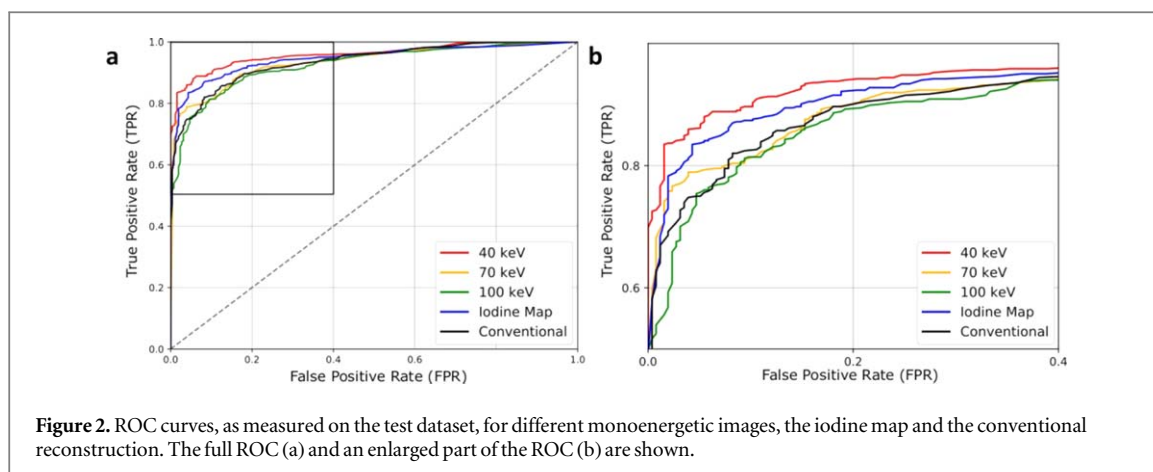
The local detection of lesions was evaluated with a distance measure. The activation map of the lesion class of each input image was up-sampled to 224×224 to match the image size using bilinear interpolation. A threshold was set to 70% of its maximum value which created one or more connected components. For each of these lesion candidates, the centre of mass was computed. Similarly, the centres of mass for all marked lesions of the ground truth annotations were calculated. The Euclidian distances

between all centres of mass for the ground truth and all centres of mass for the prediction were calculated. For the comparison of distances between the different data types, only the shortest distance was saved for each slice. The distance calculation can also be used to calculate a localisation accuracy. The distances between prediction and ground truth centres of mass were used to define TP, FP and FN. There are no TN, because there is no landmark to detect. If the distance between a prediction and a lesion is below the defined maximum distance, it is counted as TP. If a ground truth centre of mass is not within the max distance of any prediction, it is counted as FN. A prediction that is not within the max distance of any ground truth centre it is counted as FP. The accuracy is calculated per slice and evaluates the detection of all lesions. Heatmaps indicating the lesion position can be produced by using linear interpolation on the activation map and displaying it on top of the original CT image.

Results

Lesion classification

Table 1 shows a decreasing F1-score for higher energies of the VMI reconstructions as measured on the test dataset. The 40 keV VMIs produce the highest results overall. The iodine maps also yield a high F1-score, precision and recall. Conventional input images achieve F1-scores which are comparable to 70 keV. The lesion recall in table 1 quantifies how many of the true lesion slices were classified correctly on the test dataset. Training the network with 40 keV data yields the highest lesion recall. The conventional reconstruction has a noticeably lower lesion recall, which means that more lesions are misclassified. The classification performance as measured on the test dataset was also assessed with the ROC curves in figure 2. For the prediction results of the test dataset, the true positive rate is plotted over the false positive rate. A trend similar to table 1 is visible. 40 keV VMIs and iodine maps produce a higher ROC curve and larger area under the curve (AUC) than conventional images and



higher energies. The AUCs of the test datasets are 0.960 for 40 keV, 0.934 for 70 keV, 0.946 for iodine maps and 0.936 for conventional images.

Lesion localisation

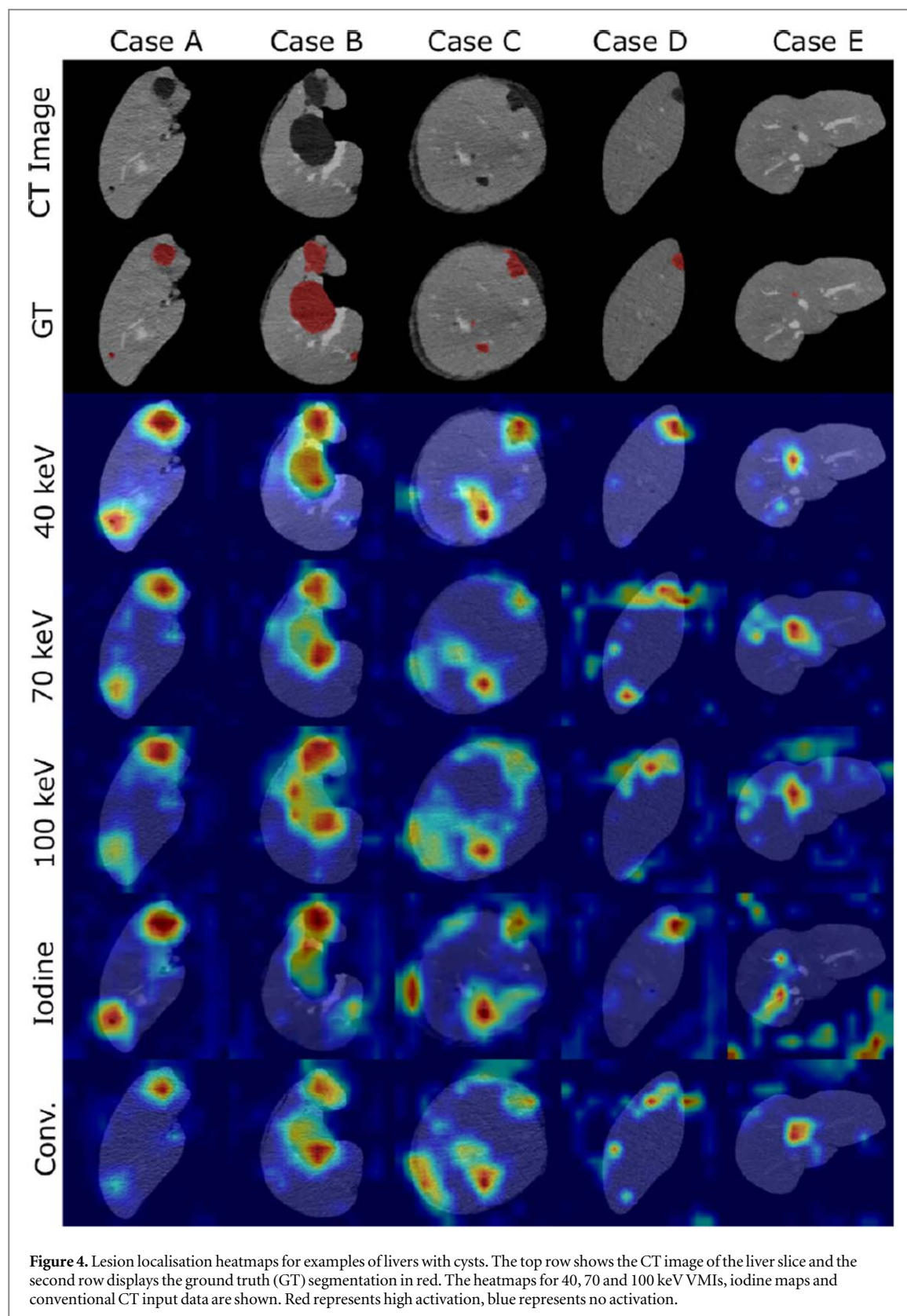
For the assessment of the localisation, only the positive cases (cases with lesion) were considered since the healthy cases do not have a specific landmark that can be detected. With conventional data distances of 10.58 ± 17.65 mm were measured; the results with iodine maps yielded the lowest distances of 8.22 ± 10.72 mm. 40 keV and 70 keV achieved distances of 8.78 ± 15.21 mm and 8.29 ± 12.97 mm, respectively, where all measurements are reported for the test dataset. All data types produced very small distances between ground truth and predicted lesion, showing that the detection of one lesion in a slice with the use of the activation maps is very accurate. With spectral data, a smaller mean distance compared to conventional data can be achieved. The localisation accuracy was calculated per slice and evaluated the detection of all lesions. Figure 3 shows the localisation accuracy for the test dataset plotted over the maximum distance measured in pixels. With an activation map of 14×14 pixels and an image size of 224×224 pixels,

one activation map pixel was scaled up to 16×16 . At a maximum distance of 16 pixels, VMIs at 40 keV yielded a localisation accuracy of 0.62 ± 0.36 , whereas the conventional data reached 0.52 ± 0.37 .

Heatmap predictions for livers with cysts and metastases are depicted in figures 4 and 5. The ground truth can be compared to the results with 40, 70 and 100 keV spectral input data, iodine maps and conventional CT input data. Big lesions were detected by all input data types. However, the 40 keV and iodine heatmaps performed best on the detection of smaller lesions. The 40 keV heatmaps were more precise and had fewer false positive results compared to the other data types. This subjective observation agrees with the evaluation of the localisation accuracy.

Region-of-interest selection

SIL with a ROI size of 48×48 achieved 90.1% correct ROIs for the lesion cases on the test dataset for 40 keV data. Iodine Maps showed 92.6% correct ROIs. Conventional data could find 83.9% correct ROIs and 70 keV achieved 88.8%. The values were higher for MIL with the same ROI size since up to three locations were used. VMIs at 40 and 70 keV found 94.2% correct

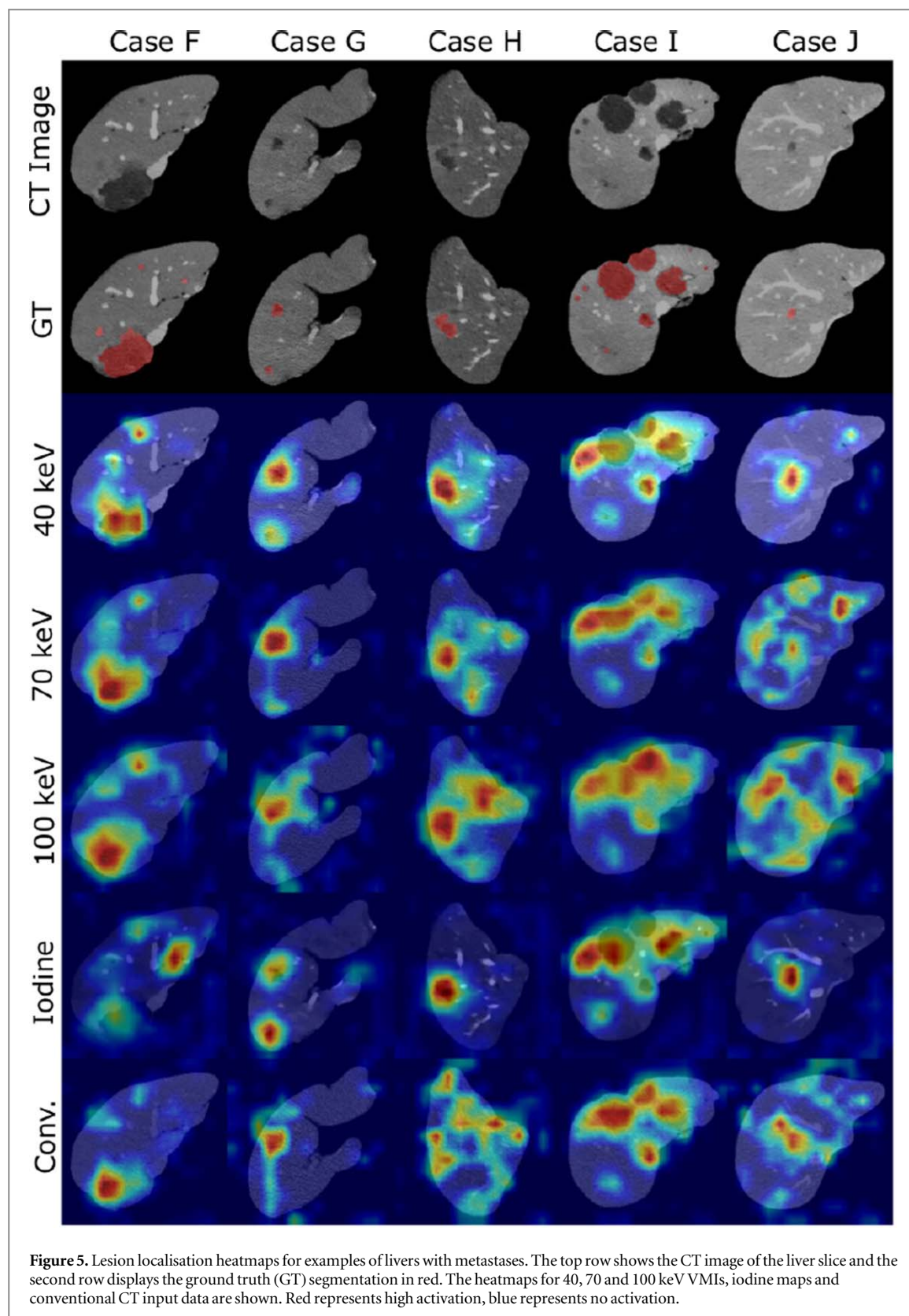


ROIs and iodine maps and conventional data produced 96.1% and 89.3%, respectively.

Lesion classification

Due to the small amount of training data, the network can reach different results. Therefore, each training was repeated five times and the mean accuracy with

standard deviation for the test dataset is reported in table 2. VMIs at 70 keV achieved the highest overall accuracy on the test set with 84.5%. The results show only small differences between the SIL and MIL approaches. The F1-score can be calculated per class similar to table 1 for the first network. The evaluation of the F1-score (table 3) and further metrics on the



individual classes demonstrate that spectral data outperforms conventional data for the lesion classes. More lesions are classified correctly compared to conventional input data.

Discussion

In this study, we illustrated the potential benefit of using the combination of spectral CT and CNNs, as opposed to CNNs and non-spectral CT, for localisation and classification of liver lesions. Both technologies have experienced increased interest over the last

Table 2. Classification Accuracy for three classes.

Input Data Type	Accuracy Mean (Median) \pm Std		
	SIL 48 \times 48 pixel	MIL 48 \times 48 pixel	MIL 70 \times 70 pixel
40 keV	0.823 (0.821) \pm 0.006	0.817 (0.815) \pm 0.008	0.817 (0.817) \pm 0.003
70 keV	0.844 (0.846) \pm 0.003	0.845 (0.848) \pm 0.015	0.845 (0.839) \pm 0.014
100 keV	0.792 (0.796) \pm 0.015	0.804 (0.805) \pm 0.009	0.800 (0.792) \pm 0.024
Iodine Map	0.695 (0.701) \pm 0.031	0.681 (0.685) \pm 0.030	0.579 (0.574) \pm 0.011
Conventional	0.803 (0.812) \pm 0.025	0.835 (0.838) \pm 0.012	0.772 (0.766) \pm 0.015

Accuracy for three class classification with ROIs. Each training was repeated 5 times. Single instance learning (SIL) and multiple instance learning (MIL) are compared for two different ROI sizes 48 \times 48 and 70 \times 70 pixels. The results are presented for different spectral and conventional input test datasets. The test dataset contained 739 ROIs (SIL) and 739 sets of three ROIs (MIL) from 33 patients.

Table 3. F1-Scores per class for three class classification.

Input Data and Method	F1-Score Mean (Median) \pm Std		
	Healthy	Cyst	Metastasis
SIL 70 keV	0.849 (0.850) \pm 0.004	0.830 (0.829) \pm 0.003	0.855 (0.853) \pm 0.006
MIL 70 keV	0.861 (0.857) \pm 0.008	0.817 (0.827) \pm 0.028	0.857 (0.861) \pm 0.017
SIL Conv.	0.843 (0.845) \pm 0.008	0.765 (0.797) \pm 0.059	0.790 (0.790) \pm 0.025
MIL Conv.	0.862 (0.859) \pm 0.006	0.807 (0.817) \pm 0.027	0.831 (0.833) \pm 0.016

F1-Score results per class for classification with ROIs using 70 keV spectral and conventional input data. Single and multiple instance learning are compared for a ROI size of 48 \times 48 pixels. All results are reported for the test dataset.

years and the combination can significantly improve the clinical workflow.

The low energy VMI reconstructions show an improved contrast between lesions and healthy tissue. In clinical routine, the optimal energy for VMIs with iodine contrast agent is between 40–70 keV [7]. A conventional image at 120 kVp has similar HU values to a VMI at 70 keV [29]. Iodine maps can show the lesions very well. Due to the lack of iodine or low iodine concentrations in cysts or hypodense metastases compared to the surrounding liver tissue, these lesions appear as hypodense areas. The classification results for the weakly-supervised CNN illustrate the advantage of the low energy VMIs over the other input types. The CNN can benefit from the higher contrast between lesions and healthy tissue, which means more slices with lesions are detected correctly for lower VMI energies. The F1-score is higher, the lower the VMI energy is (see table 1). The performance of the conventional images is similar to a VMI at 70 keV, which can be explained by the similar HU values in both images. When comparing the localisation results, the same trend can be seen for the localisation accuracy (figure 3).

The heatmaps presented in figures 4 and 5 visually illustrate the benefit of the 40 keV spectral data for the task of lesion localisation. Big lesions are detected by all input data types. However, the 40 keV heatmaps outperform higher energy and conventional heatmaps for the detection of smaller lesions. They are more precise, have a higher response for small lesions and fewer

false positive detections. The iodine maps also perform well for the detection of small lesions, but sometimes give a lower response for the biggest lesion of the slice, which makes VMIs at 40 keV overall more attractive for hypodense lesion detection.

A common limitation for the application of deep learning in the medical domain is the need for radiologists to select ROIs manually. In the current work, an automatic workflow omitting the need for expert interaction was presented. The lesions are classified with ROIs chosen automatically based on the activation maps of the weakly-supervised CNN. The quality of the ROIs was assessed; here, a good performance of low energy spectral data for the lesion localisation can be confirmed. For VMIs at 40 keV and 70 keV, more accurate ROIs are selected compared to 100 keV and the conventional CT input. The iodine maps yield an even higher number of correct ROIs. Over 90% of all selected ROIs for 40 keV VMIs were correct, showing that this is a relatively robust and reliable way to employ spectral CT data when an expert ROI selection is not available.

The comparison of the different input data types for the second network (classification into three classes - healthy, cyst, metastasis) used five repetitions of each training, subsequently the mean values were calculated (tables 2 and 3). The lesion classification showed sound results for both methods (SIL and MIL) tested in this study. Interestingly, the iodine maps performed worst in the three class classification, despite having selected the best ROIs. In the iodine map,

lesions appear as dark shapes with strong contrast to the surrounding tissue. Lesions in general can be detected well by the network (as seen in table 1 and figures 4 and 5), but the examples for cyst and metastasis iodine maps look very similar and it is harder to distinguish between the lesion types on this dataset. In future studies, it will be necessary to evaluate if the performance improves for hyperdense lesions. The overall accuracy was best for VMIs at 70 keV. The classification into the three classes 'healthy', 'cyst' and 'metastasis' can be performed with 84% accuracy despite the presence of some wrong ROIs in the training and test set.

This study had some limitations. Above all, the study included only a small sample size which was obtained from scans of 172 different patients and accommodated all three target classes (hypodense metastases, cyst and healthy) and was further split into the training, validation and testing tests. In addition, manual lesion segmentations, which were performed by an experienced radiologist and served as a ground truth, might include minor segmentation errors which could slightly affect the evaluation results. However, these stated limitations exemplify the importance of automatic algorithms such as a weakly-supervised CNN for advancements in CAD technology. In addition, the CNN algorithms and architecture that were used in this study for the classification and localization of lesions are not state-of-the-art techniques but rather standard CNNs that were utilized to demonstrate that existing computer-vision methods can indeed benefit from the additional information present in spectral images. It is also important to note that the level of improvement may depend on the specific utilized technique and may defer for different algorithms. A comparison of state-of-the-art techniques to CNNs which take advantage of spectral information is within the scope of a future study. Moreover, the effects of reduced radiation dose levels and their associated increase in image noise were not evaluated as part of this study. Since these effects are expected to decrease the resulting accuracies of both the conventional and the spectral datasets, further study is required to evaluate whether or not the demonstrated benefit of utilizing spectral information is maintained at low and ultra-low radiation dose levels. Finally, both the training and test datasets used in this study did not include image artifacts such as artifacts caused by metals in the vicinity of the liver. While dual-energy CT is known for its reduced metal artifacts compared to conventional CT [30], further investigations are required in order to assess the impact of metal and other image quality artifacts on the interplay between modern CNNs and spectral CT technology.

For future work investigating other pathologies (e.g. hyperdense lesions, hepatocellular carcinoma) and applying the lesion localisation to challenging cases such as fatty livers will be of interest. Additionally, spectral extensions to the algorithm studied in

this work, such as utilizing multiple spectral results as multi-parametric input or an automatic keV selection for maximizing the accuracy of specific algorithmic tasks, i.e. classification versus localization, have the potential to increase the overall accuracy of the method. Finally, considering all contrast phases (native, arterial, and portal venous) as input into a CNN could further improve the diagnostic accuracy. To achieve this goal, next generation spectral CT systems, which are equipped with photon-counting detectors, are necessary because aligned data across contrast agent phases can be generated [31–33]. In summary, the current possibilities presented in this study as well as further developments of both technologies have the potential to significantly aid the diagnostic decision process.

Summary statement

The combination of spectral CT and CNNs has the potential to improve the detection and classification of small liver lesions.

ORCID iDs

Nadav Shapira  <https://orcid.org/0000-0002-9294-7001>

Felix K Kopp  <https://orcid.org/0000-0003-2069-4444>

Gregor Pahn  <https://orcid.org/0000-0002-5169-0158>

Alexander A Fingerle  <https://orcid.org/0000-0002-6602-1360>

References

- [1] American Cancer Society 2015 *Global Cancer Facts & Figures* 3rd edn (Atlanta: American Cancer Society)
- [2] Belghiti J, Cauchy F, Paradis V and Vilgrain V 2014 Diagnosis and management of solid benign liver lesions *Nature Reviews Gastroenterology & Hepatology* **11** 737–49
- [3] Cogley J R and Miller F H 2014 MR imaging of benign focal liver lesions *Radiologic Clinics* **52** 657–82
- [4] Feldman M, Friedman L S and Brandt L J 2010 *Sleisenger and Fordtran's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management 2* (Philadelphia: Elsevier Health Sciences)
- [5] Saba L and Suri J S 2014 *Multi-Detector CT Imaging: Abdomen, Pelvis, and CAD Applications 2* (Boca Raton: CRC Press)
- [6] Johnson T et al (ed) 2011 *Dual Energy CT in Clinical Practice* (Berlin: Springer Science & Business Media)
- [7] McCollough C H, Leng S, Yu L and Fletcher J G 2015 Dual-and multi-energy CT: principles, technical approaches, and clinical applications *Radiology* **276** 637–53
- [8] Flohr T et al 2006 First performance evaluation of a dual-source CT (DSCT) system *European radiology* **16** 256–68
- [9] Zhang D, Li X and Liu B 2011 Objective characterization of GE discovery CT750 HD scanner: gemstone spectral imaging mode *Med. Phys.* **38** 1178–88
- [10] Carmi R, Naveh G and Altman A 2005 Material separation with dual-layer CT *Nuclear Science Symposium Conference Record, 2005 IEEE 4* (IEEE) 3

- [11] Ehn S *et al* 2018 Assessment of quantification accuracy and image quality of a full-body dual-layer spectral CT system *Journal of applied clinical medical physics* **19** 204–17
- [12] Primak A N *et al* 2007 Noninvasive differentiation of uric acid versus non-uric acid kidney stones using dual-energy CT *Acad. Radiol.* **14** 1441–7
- [13] Wang Q, Shi G, Qi X, Fan X and Wang L 2014 Quantitative analysis of the dual-energy CT virtual spectral curve for focal liver lesions characterization *Eur. J. Radiol.* **83** 1759–64
- [14] Agrawal M D *et al* 2014 Oncologic applications of dual-energy CT in the abdomen *Radiographics* **34** 589–612
- [15] Yu L *et al* 2011 Virtual monochromatic imaging in dual-source dual-energy CT: radiation dose and image quality *Med. Phys.* **38** 6371–9
- [16] Rassouli N *et al* 2017 Detector-based spectral CT with a novel dual-layer technology - principles and applications *Insights Imaging* **8** 589–98
- [17] SELLERER T *et al* 2018 Dual-energy CT: a phantom comparison of different platforms for abdominal imaging *European Radiology* **28** 2745–55
- [18] Gletsos M *et al* 2003 A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier *IEEE Trans. Inf. Technol. Biomed.* **7** 153–62
- [19] Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *Advances in Neural Information Processing Systems* **60** 1097–105
- [20] Yasaka Y, Akai H, Abe O and Kiryu S 2018 Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: A preliminary study *Radiology* **286** 887–96
- [21] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. of the IEEE conference on computer vision and pattern recognition* 770–8
- [22] Oquab M, Bottou L, Laptev I and Sivic J 2015 Is object localization for free?—Weakly-supervised learning with convolutional neural networks *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* 685–94
- [23] Zhou B, Khosla A, Lapedriza A, Oliva A and Torralba A 2016 Learning deep features for discriminative localization *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conf.* 2921–9
- [24] Hwang S and Kim H E 2016 Self-transfer learning for weakly supervised lesion localization *International Conference on Medical Image Computing and Computer-Assisted Intervention (Cham: Springer)* 239–46
- [25] Wang X *et al* 2017 Chestxray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE)* 3462–71
- [26] Santamaria-Pang A *et al* 2010 Automated liver lesion characterization using fast kVp switching dual energy computed tomography imaging *Medical Imaging 2010: Computer-Aided Diagnosis* 7624 (International Society for Optics and Photonics)
- [27] Gauriau R *et al* 2015 Multiple template deformation application to abdominal organ segmentation 2015 *IEEE 12th International Symposium on Biomedical Imaging (ISBI) (IEEE)* 359–62
- [28] Christ P F *et al* 2017 Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks arXiv:1702.05970
- [29] Rassouli N *et al* 2017 Assessment of 70-keV virtual monoenergetic spectral images in abdominal CT imaging: A comparison study to conventional polychromatic 120-kVp images *Abdominal Radiology* **42** 2579–86
- [30] Neuhaus V *et al* 2017 Metal artifact reduction by dual-layer computed tomography using virtual monoenergetic images *Eur. J. Radiol.* **93** 143–8
- [31] Muenzel D *et al* 2016 Spectral photon-counting CT: initial experience with dual-contrast agent K-edge colonography *Radiology* **283** 723–8
- [32] Muenzel D *et al* 2017 Simultaneous dual-contrast multi-phase liver imaging using spectral photon-counting computed tomography: a proof-of-concept study *European Radiology Experimental* **1** 25
- [33] Symons R *et al* 2017 Photon-counting CT for simultaneous imaging of multiple contrast agents in the abdomen: an *in vivo* study *Med. Phys.* **44** 5120–7