Liver segmentation and metastases detection in MR images using convolutional neural networks

Citation for published version (APA):

DOI:
10.1117/1.JMI.6.4.044003

Document status and date:
Published: 01/10/2019

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Liver segmentation and metastases detection in MR images using convolutional neural networks

Mariëlle J. A. Jansen
Hugo J. Kuijf
Maarten Niekel
Wouter B. Veldhuis
Frank J. Wessels
Max A. Viergever
Josien P. W. Pluim

Liver segmentation and metastases detection in MR images using convolutional neural networks

Mariëlle J. A. Jansen, a,* Hugo J. Kuijf, a Maarten Niekel, b Wouter B. Veldhuis, b Frank J. Wessels, b Max A. Viergever, a and Josien P. W. Pluim a,b

UMC Utrecht and Utrecht University, Image Sciences Institute Utrecht, The Netherlands
UMC Utrecht, Department of Radiology Utrecht, The Netherlands

Abstract. Primary tumors have a high likelihood of developing metastases in the liver, and early detection of these metastases is crucial for patient outcome. We propose a method based on convolutional neural networks to detect liver metastases. First, the liver is automatically segmented using the six phases of abdominal dynamic contrast-enhanced (DCE) MR images. Next, DCE-MR and diffusion weighted MR images are used for metastases detection within the liver mask. The liver segmentations have a median Dice similarity coefficient of 0.95 compared with manual annotations. The metastases detection method has a sensitivity of 99.8% with a median of two false positives per image. The combination of the two MR sequences in a dual pathway network is proven valuable for the detection of liver metastases. In conclusion, a high quality liver segmentation can be obtained in which we can successfully detect liver metastases. © 2019 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.6.4.044003]

Keywords: dynamic contrast-enhanced MRI; diffusion weighted MRI; liver; segmentation; detection; deep learning.

Paper 19077RRR received Mar. 20, 2019; accepted for publication Sep. 17, 2019; published online Oct. 15, 2019.

1 Introduction

Primary tumors, such as neuroendocrine and colorectal tumors, have a high likelihood of developing metastases in the liver.1,2 Early detection of (new) liver metastases is crucial since it improves patient outcome.3–5 To follow disease progress, radiologists check for tumor growth and new (liver) metastases in computed tomography (CT) or magnetic resonance (MR) images.

While CT has long been the modality of choice in detecting and monitoring liver tumors, MRI has gained interest due to a better lesion-to-liver contrast and because it does not use ionizing radiation.6,7 Dynamic contrast-enhanced (DCE) MR images have a high sensitivity and specificity for visual detection of liver metastases. Moreover, the combination of DCE-MR and diffusion weighted (DW) MR images turns out to be even more effective in the visual detection of liver metastases and visual censoring of mimics.8–10

The automatic detection and characterization of liver metastases remains a challenging task, given the heterogeneous appearance of liver metastases on MR images.11,12 Radiologists have a liver lesion detection rate between 87% and 95%, when using both DCE-MR and DW-MR images.7,9 Automatic liver metastases detection could aid radiologists in finding metastases more efficiently and effectively.

The liver metastases detection method we propose is a two-step method: a liver segmentation step followed by a lesion detection step within the segmented liver. Most liver segmentation methods published have been developed for CT, with best results by convolutional neural network (CNN)-based methods.13,14 Fewer methods have been developed for segmentation of the liver in MR images. Those known are primarily based on watershed,15 active contouring,16 atlases,17 or shape models.18

For automatic detection of liver lesions, several methods have been proposed, all based on CT images.14,20,21–23 MR data were employed by a few methods to detect hepatocellular carcinomas (either DCE-MRI24 or DW-MRI25), but not for metastases.

Our aim is to develop and evaluate a two-step liver metastases detection method for MR images, based on fully convolutional neural networks (FCN). In the first step, a liver segmentation method utilizing the dynamic nature of the DCE-MR images is presented, based on previous work.20 This is followed by a method for the detection of metastases within the liver region using both DCE-MR and DW-MR images as input.

2 Data

The study comprises MR data of 121 patients with a clinical focus on the liver from the University Medical Center (UMC) Utrecht, The Netherlands, acquired between February 2015 and February 2018. The UMCU Medical Ethical Committee has reviewed this study and informed consent was waived due to its retrospective nature.

All patients underwent a clinical MR examination, including a DCE-MR series and DW-MRI. The DCE-MR series was acquired in six breath holds with one to five three-dimensional (3-D) images per breath hold, with the following parameters: TE: 2.143 ms; TR: 4.524 ms; flip angle: 10 deg. After acquiring the first image, gadobutrol (0.1 ml/kg Gadovist of 1.0 mmol/ml at 1 ml/s) was administered at once, followed by 25 ml saline solution at 1 ml/s. In total, 16 3-D images per patient were acquired with 100 slices and matrix sizes of 256 × 256. Voxel size was 1.543 mm × 1.543 mm × 2 mm.

The DW-MR images were acquired with three b-values (10, 150, and 1000 s/mm²) using a protocol with the following parameters: TE: 70 ms; TR: 1.660 ms; and flip angle, 90 deg.
Each $b$-value image was acquired with 42 slices and matrix sizes of $256 \times 256$. Voxel size was $1.758 \text{ mm} \times 1.758 \text{ mm} \times 5 \text{ mm}$.

2.1 Preprocessing

2.1.1 DCE-MR

All DCE-MR data sets were corrected for motion using a groupwise registration.\(^{27}\) The groupwise registration method registers all images simultaneously to a common space by minimizing a cost function based on principle component analysis and applying a B-spline transformation. The registration is applied on four resolutions with 500 iterations each. After registration, a zero-mean-unit-variance rescaling was applied to all intensity values between the 0th and 99.8th percentile of the intensity histogram of the DCE-MR series. The 99.8th percentile intensity was assumed to correspond to the contrast agent peak in the aorta.

The images were combined per phase by averaging the fourth dimension. The series started with the precontrast image, followed by the other phases: the early arterial phase, the late arterial phase, the hepatic/portal-venous phase, the late portal-venous/equilibrium phase, and the late equilibrium phase. The six phases were used as input images for the FCN.

2.1.2 DW-MR

The DW-MR images were non-linearly registered to the DCE-MR space using elastix, a toolbox for linear and nonlinear registration of medical images.\(^{28}\) The fixed image was the mean DCE-MR, obtained by averaging over the six phases. First, a rigid transformation was applied on two resolutions with 2000 iterations each, followed by a b-spline transformation on one resolution with 1000 iterations and a grid spacing of $60 \times 60 \times 40 \text{ mm}$. Normalized mutual information was used as metric (The parameter files used are available at Ref.\(^29\)). A mask was used to focus the registration on the liver. The mask was obtained from the automatic liver segmentation (Sec. 3.1), morphologically dilated with a $10 \times 10 \times 10$ structuring element. A substantial dilation was chosen to make sure the boundary of the liver was included.

Ten out of the 121 MRI examinations were excluded from the study because of a failed registration of the DW-MRI on the DCE-MR series. The exclusion was based on visual evaluation.

The intensities of the registered DW-MR data set were also normalized with a zero-mean-unit-variance rescaling between the 0th and 99.8th percentile of the intensity histogram of the DW-MRI.

2.2 Annotations

The liver was annotated in 55 DCE-MR series, of which only 16 series contained metastases that were segmented. In addition, we included another 56 DCE-MR series in which the metastases were segmented, resulting in a total of 72 DCE-MR series with annotated metastases.

2.2.1 Liver segmentation

Fifty-five DCE MR series were used for liver segmentation. The data sets were randomly divided into 33 data sets for training, 3 for validation, and 19 data sets for testing. The validation set was used for tuning the hyperparameters and the evaluation of the CNN model during method development. The test set was used to evaluate the final CNN model in an independent manner.

The liver was manually contoured by two observers; see Fig. 3 for segmentation examples. The first observer annotated the training, validation, and test sets. The annotation of the test set is indicated as O1.1. The first observer repeated the annotations of the test set at least 1 week later (O1.2). The second observer annotated the test set once (O2). This was done to estimate the inter- and intraobserver agreement. A radiologist with more than 10 years of experience in liver MR analysis verified all manual annotations and provided corrections where needed. Liver lesions were included in the annotation, and this network was therefore trained to recognize liver lesions as liver tissue. The first set of annotations of the first observer was used as reference in the experiments.

2.2.2 Liver metastases detection

Seventy-two MR data sets were used for the liver metastases detection. The data set included mainly colorectal metastases, neuroendocrine metastases, and some other metastases types (i.e., other gastrointestinal metastases and breast metastases). The data sets were randomly divided into 55 data sets for training ($n = 50$) and validation ($n = 5$), and 17 data sets for testing. The training and validation sets contained 334 metastases in total, with on average six metastases per liver (range: 1 to 31 metastases). The test set contained 86 metastases in total, with on average five metastases per liver (range: 1 to 32 metastases).

The metastases were manually annotated on the DCE-MR images by a radiologist in training and were verified by a radiologist with more than 10 years of experience.

3 Methods

3.1 Liver Segmentation

An FCN\(^{30}\) with dilated convolutions was implemented, for which the six phases of the DCE-MR images were the channels of the input image. The dilated FCN consisted of nine convolutional layers in total. The first seven layers had a $3 \times 3$ convolution and 32 kernels. The dilation rates ranged from 1 to 16. The final two layers had a $1 \times 1$ convolution. ReLU activation and batch normalization were used in all the convolutional layers, except for the final layer, which had a softmax activation. A dropout layer was applied between the eighth and the ninth layers with a dropout rate of 0.5. This network had a receptive field of $67 \times 67$ pixels. Figure 1 gives an overview of the network architecture. In our previous work on liver segmentation,\(^{26}\) we also considered the popular U-net architecture. However, this architecture had a slightly worse performance than the FCN architecture used here.

The loss was calculated by a similarity metric based on the Dice similarity coefficient (DSC): $\frac{2|X \cap Y|}{|X|+|Y|}$, where $X$ is the predicted segmentation, $Y$ is the ground truth mask, and $s$ is a small number to prevent dividing by zero (set to 1e-5).\(^{31}\) Glorot uniform\(^{32}\) was used as the initializer and Adam as the optimizer with a learning rate of 0.001. The network was trained for 100,000 iterations, with six images per minibatch. The total number of two-dimensional (2-D) slices for training was 3300 slices. No data augmentation was applied.

The data were processed by the network per 2-D slice consisting of six channels. For the evaluation of the liver segmentation, the probability output was postprocessed to a binary image. The threshold of 0.5 was applied to the probability output of the network. Then, 3-D hole filling was performed, so all
holes caused by liver lesions were filled. To remove small spurious segmentations, the largest connected component was selected as the final segmentation.

3.2 Liver Metastases Detection

A dual-pathway FCN was implemented, for which the input images of one path were the six DCE-MR phase images and for the other path the three DW-MR images. The six (or three) 2-D images were combined to one input image, with six (or three) channels. Each pathway had 13 convolutional layers with a $3 \times 3$ convolution and 64 kernels, split in five blocks with different dilation rates, ranging from 1 to 8. The feature maps at the end of each block of each pathway were concatenated in the third dimension, resulting in a feature map with 640 kernels, and were passed to two convolutional layers with a $1 \times 1$ convolution with 128 and 2 kernels, respectively. This resulted in a receptive field of $123 \times 123$ pixels. The individual pathways were inspired by the P-net architecture. Figure 2 gives an overview of the network architecture.

In addition, the network was also trained and evaluated as a single pathway FCN, once with only the DCE-MR images as input images with six channels and once with the DCE-MR and DW-MR images concatenated as input images with nine channels. In this manner, the additional value of the DW-MR images and the addition of a second pathway in the detection method can be determined. The single pathway FCN was identical to the top pathway in Fig. 2. In the concatenation layer, the feature maps at the end of each block were concatenated in the third dimension.

Categorical cross-entropy was used as the loss function. ReLU activation and batch normalization were used in all the convolutional layers, except for the final layer, which had a softmax activation. Two dropout layers were applied before and after the second-to-last layer. The dropout rate was set to 0.2. The classes were weighted based on class frequencies. He uniform was used as the initializer and Adam as the optimizer with a learning rate of 0.0001. The network was trained for 10,000 iterations, with four images per mini batch. The slices to train on were limited to the slices containing lesions for a more balanced data set, resulting in a total of 1619 2-D slices for training and 180 slices for validation.

Twenty-five patches of $128 \times 128$ pixels were taken from each slice for data augmentation. The patches originated from the liver region and have overlapping areas. Data augmentation was applied by random rotation of the patches, with rotation angles of ±45°.

The data of the test set were processed by the network per 2-D image slice. The input image slice consisted either of six channels of the DCE-MR phase images, three channels of DWI images, or nine channels, depending on the network. The probability output was masked by the liver segmentation, which was dilated with a $5 \times 5$ structuring element. The dilation of the liver segmentation is a safety measure to ensure that small failures in the liver segmentation would not lead to undetected liver metastases.
For the evaluation of the detection, the masked probability output was postprocessed to a binary image. All pixels with a probability output higher than the threshold of 0.5 were labeled as metastasis. Morphological closing with a structuring element of $3 \times 3 \times 3$ was applied to fill any holes in the binary image. The morphological closing was followed by a morphological opening with a plus-shaped structuring element of $3 \times 3$, to remove any remaining noise in the detection results. The resulting binary image was divided into separate objects representing individual metastases, using voxel clustering with 26-neighborhood connection.

4 Experiments

4.1 Liver Segmentation

The performance of the liver segmentation was evaluated based on the DSC, the relative volume difference (RVD), and the modified Hausdorff distance (HD) at the 95th percentile. The first annotations of the first observer of the test set (O1.1) are used to evaluate the automatic segmentation since these were made by the same observer in the same session as the annotations of the training and validation sets:

\[
\text{DSC}: \frac{2X \cap Y}{X + Y},
\]

where $X$ is the automatic segmentation and $Y$ is the annotation O1.1:

\[
\text{RVD}: \frac{X - Y}{Y} \times 100\%
\]

HD at 95th percentile: $\max[h_{05}(X, Y), h_{05}(Y, X)]$, with $h_{05}(X, Y) = K_{05}^{\text{seg}} \min \{\epsilon \in \mathbb{R} | |y - x|\}$, Here, $X$ is the set of boundary points $\{x_1, \ldots, x_n\}$ of the automatic segmentation result, and $Y$ is the set of boundary points $\{y_1, \ldots, y_n\}$ of the annotation O1.1. $K_{05}^{\text{seg}}$ denotes the 95th ranked value in the set of distances between all boundary points in $X$ and the closest boundary points in $Y$.

These three metrics were computed on the predicted segmentation results. In addition, the three metrics were computed for the second annotations of the first observer (O1.2) and the annotations of the second observer (O2) to obtain the intra- and interobserver agreement, respectively.

4.2 Liver Metastases Detection

A liver metastasis was considered detected, and thus a true positive object, when the manual annotation and the predicted segmentations had an overlap greater than 0.

For the evaluation of the liver metastases detection, the true positive rate (TPR) and the number of false positives per case (FPC) were reported. The TPR was calculated as the number of true positive objects divided by the total number of true lesion objects. The FPC was calculated as the number of objects not overlapping with any true metastasis object. In addition, the TPR and FPC were given for several thresholds in a free-response receiver operating characteristic (FROC) curve.

The same metrics were applied to the segmentation results of the single pathway networks.

An expert radiologist verified the results and determined the underlying physiology of selected false positive objects.

5 Results

Figure 3 shows some examples of the late arterial phase of the DCE-MRI, the DW-MRI with b-value 150 s/mm², the manually annotated liver and metastases, and the automatic segmentation of the liver with the detection of the liver metastases involving both DCE-MR and DW-MR images in the dual pathway FCN. It shows good liver and lesion segmentations, with a false positive object in the last row, which is verified to be a cyst.

5.1 Liver Segmentation

The median DSC is 0.95 for the automatic segmentation, 0.94 for the interobserver agreement, and 0.96 for the intraobserver agreement. The median RVD is $-1.6\%$ for the automatic segmentation, $-3.4\%$ for the interobserver agreement, and $-1.5\%$ for the intraobserver agreement. The median HD is 5.5 mm for the automatic segmentation, 5.6 mm for the interobserver agreement, and 3.1 mm for the intraobserver agreement. The distributions of the DSC, RVD, and modified HD are given in boxplots in Fig. 4.

In the boxplots, annotations O1.1 are used to evaluate the automatic liver segmentations. Some variance is present between the different annotations of the observers; therefore, we also compared the automatic method with all manual annotations. The median [interquartile range (IQR)] values of the DSC, RVD, and modified HD for the automatic segmentation, the interobserver agreements, and the intraobserver agreement using the three manual annotations are given in Table 1. The upper three rows give the same results as the boxplots and the lower three rows give additional results.

Example results of the automatic liver segmentation are shown in the last column of Fig. 3 and in Fig. 8.

5.2 Liver Metastases Detection

Figure 5 shows the TPR and the FPC for all subjects. The average TPR for the single pathway network using only DCE-MRI is 0.645 and for using both DCE-MRI and DW-MRI is 0.722. The average TPR is 0.998 for the dual pathway network using both DCE-MRI and DW-MRI. The median FPC is five for the single pathway network using only DCE-MRI, six when using both MR sequences, and two for the dual pathway network. The single pathway has 111 and 146 false positive objects in total, using only DCE-MRI and both MR sequences, respectively, while the dual pathway has 59 false positive objects in total.

Figure 6 shows the number of detected metastases relative to the total number of metastases per size category. Both of the single pathway networks fail on the smaller metastases in particular.

Figure 7 shows the FROC curve of the mean TPR versus the median FPC for different thresholds $T$ applied to the output of both networks. The thresholds range from 0.90 to 0.00 with steps of 0.10. At the highest threshold ($T = 0.90$), only 42% of the metastases were detected with a median FPC of one (in total, 31 false positive objects are present) using only the DCE-MRI images, and 65% of the metastases were detected with a median FPC of one (in total, 25 false positive objects are present) using both MR sequences in the single pathway. For the dual pathway, 96% of the metastases were detected, with a median FPC of one (in total, 19 false positive objects are present). At the lowest threshold ($T = 0.0$), 84% were detected with only DCE-MRI images, 91% were detected with both MR...
sequences in the single pathway, and all the lesions were detected with the dual pathway network, but all at the cost of many false positives.

Figure 8 shows five examples of the lesion detection results, including the liver segmentation result. The green pixels are true positive, blue pixels are false positive, and red pixels are false negative. Note that true positive objects are connected components that overlap with a manually annotated metastasis. A true positive object can nonetheless have some undersegmentation (red pixels, example indicated with arrow head) or oversegmentation (blue pixels, example indicated with arrow).

Fig. 3 Three examples of liver and metastases segmentations. From left to right, each column represents a late arterial phase DCE-MR image, the registered DW-MR image with $b$-value 150 s/mm$^2$, the manually annotated liver (orange) and metastases (green), and the automatically segmented liver (orange) and metastases (green). The bottom row shows a false positive object in the anterior side of the liver for the automatic detection, which is a cyst.
6 Discussion

6.1 Liver Segmentation

Our proposed liver segmentation method is able to provide high quality liver segmentations using DCE-MR images. The method is able to deal with common difficulties in MR segmentation as well as the presence of lesions.

The automatic method performs well on the test set of annotation O1.1. The results of the automatic method are similar to the interobserver results, as can be seen in the top rows of Table 1. Observer 1 annotated the training set, so it is not surprising that the automatic segmentations do well on the test set, which was annotated by the same observer in the same annotation session. There is some intraobserver variance between O1.1 and O1.2, and the second annotations appear systematically smaller than the first annotations since the RVD values involving O1.1 are mostly negative. Still, the automatic method also performs well when evaluated with annotations O1.2, as can be seen in the bottom rows of Table 1. Annotations O2 were used to calculate the interobserver variance and were additionally used to evaluate the automatic method with another observer to verify whether the method generalizes well. The results using O2 for evaluation are still good, although slightly worse than the automatic method evaluated with O1. This is as expected because the first observer annotated the training and validation sets.

The automatic liver segmentation encountered few problems. However, an outlier in all three metrics was found in one case, where a large lesion (337 ml) occurred at the edge of the liver. This outlier has an RVD of $-15\%$, as a result of undersegmentation of the liver in this lesion area. Even though the lesion is bigger than the receptive field of the neural network, the method accomplished recognizing half of the lesion tissue as part of the liver. However, large lesions at the edge of the liver are a potential pitfall for the proposed liver segmentation method.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>DSC</th>
<th>RVD (%)</th>
<th>HD (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic – O1.1</td>
<td>0.95</td>
<td>-1.6</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>[0.93 to 0.96]</td>
<td>[-3.5 to 0.8]</td>
<td>[4.0 to 8.0]</td>
</tr>
<tr>
<td>Interobserver</td>
<td>0.94</td>
<td>-3.4</td>
<td>5.6</td>
</tr>
<tr>
<td>O2 – O1.1</td>
<td>[0.93 to 0.95]</td>
<td>[-5.1 to 0.6]</td>
<td>[4.3 to 6.8]</td>
</tr>
<tr>
<td>Intraobserver</td>
<td>0.96</td>
<td>-1.5</td>
<td>3.1</td>
</tr>
<tr>
<td>O1.2 – O1.1</td>
<td>[0.96 to 0.97]</td>
<td>[-1.9 to -0.9]</td>
<td>[2.5 to 3.1]</td>
</tr>
<tr>
<td>Automatic – O1.2</td>
<td>0.95</td>
<td>-0.1</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>[0.93 to 0.96]</td>
<td>[-2.8 to 2.2]</td>
<td>[3.7 to 8.0]</td>
</tr>
<tr>
<td>Automatic – O2</td>
<td>0.95</td>
<td>0.5</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>[0.93 to 0.95]</td>
<td>[-3.8 to 5.5]</td>
<td>[4.9 to 9.6]</td>
</tr>
<tr>
<td>Interobserver</td>
<td>0.95</td>
<td>-2.5</td>
<td>4.4</td>
</tr>
<tr>
<td>O2 – O1.2</td>
<td>[0.94 to 0.96]</td>
<td>[-4.0 to 1.6]</td>
<td>[3.7 to 4.8]</td>
</tr>
</tbody>
</table>

Fig. 5 The TPR and the number of false positives per subject for the single pathway and the dual pathway networks.

Journal of Medical Imaging 044003-6 Oct–Dec 2019 © Vol. 6(4)

Jansen et al.: Liver segmentation and metastases detection in MR images...
It is hard to compare the liver segmentation results with other methods since the data sets and MRI sequences used are not the same. A rough comparison with recently proposed automatic liver MR segmentation algorithms\(^\text{17,18,25}\) can be made by considering the reported values for the DSC. For the mentioned studies, the DSC ranged from 0.87 to 0.91, while we report a DSC of 0.95.

### 6.2 Liver Metastases Detection

The detection of all liver metastases is of great importance for adequate treatment.\(^\text{36}\) The detection method with only DCE-MR images has an average TPR of 0.645 and a median of 5 FPC. The detection method with the single pathway using both MR sequences is more effective, with an average TPR of 0.722 and a median of six FPC. The dual pathway detection method is able to detect almost all metastases at a threshold level of 0.50 with a median of two FPC. This shows the importance of adding DW-MR images to the lesion detection method and processing the two MR sequences in separate pathways. DW-MR images are highly sensitive for lesions and, in combination with the DCE-MR images, the method is able to detect metastases at a high detection rate and a low number of false positives. Figure 8 shows some visual examples of the differences in detection of these three networks.

The usage of the two MR sequences requires the registration of the DW-MR images to the DCE-MR image space. Ten registrations failed, which led to the exclusion of these images from the study. This is likely caused by the free-breathing protocol used for the DW-MR images, which may result in deformations that cannot be corrected by the registration method. A more dedicated registration approach or a breath-hold DW-MR acquisition may improve the robustness of the method.

One small metastasis was missed by the dual pathway network. The metastasis is only five pixels in size and was detected by the dual pathway with a probability higher than 0.50. However, it was deleted during the binary opening in the postprocessing step, where it was seen as noise. This reduction of noisy false positives comes at the cost of missing very small metastases.

The missed metastases by the single pathway networks are mostly smaller than \(2\text{ cm}^3\), as can be seen in Fig. 6. These smaller metastases may differ in appearance from larger ones because certain features, such as rim enhancement, cannot be expressed in only a few voxels of the DCE-MR images. The network with only DCE-MR input images is uncertain about these metastases, which leads to low outcome probabilities. However, the single pathway with both DCE-MR and DWI-MR is also less able to detect these smaller metastases. The network seems to fail to create adequate feature maps to detect small metastases, when it combines the two MR sequences in the first convolutional layer. The dual pathway network is allowed to create feature maps for the two MR sequences separately. The phases of the DCE-MR images are correlated over time, as are the three \(b\)-value images of the DW-MR. Both MR sequences contain useful characteristics. In the dual pathway network, the feature maps are solely composed of the MR sequence presented to that pathway, representing those characteristics, and the feature maps are combined at the end. This might

---

**Fig. 6** The number of lesions plotted against the size of the lesions. The colored part represents the detected lesions and the white part the missed lesions at a threshold of 0.50.

**Fig. 7** FROC curve of the mean TPR and the median number of FPC for threshold \(T\) ranging from 0.90 to 0.0 in steps of 0.10. The circles represent the TPR and FPC at threshold \(T = 0.50\).
explain the difference in the detection of the small metastases of the two networks. In addition, the dual pathway has two times more parameters as the single pathway. Figure 8, first column, shows some examples of subcentimeter-sized metastases.

The dual pathway network detects almost all metastases, but also incorrectly marks other objects as metastases. Most of these false positives are caused by objects that are not or scarcely represented in the training set and thus have an appearance unknown to the network. These unknown appearances could be other lesions or liver conditions such as cirrhosis, or an inhomogeneous fat distribution in the liver. The network has mostly seen healthy liver parenchyma with metastases during training and has learned to distinguish these, but it is uncertain about conditions and lesions not seen during training. Figures 3 (last row) and 8 (third column) show examples of cysts marked as a metastasis by the detection method.

Inspection of the 19 false positive objects for a threshold level of 0.90 in the postprocessing step reveals that 10 of the false positives are lesions: nine cysts and one hemangioma. Three of the false positives are blood vessels and one false positive is because of an inhomogeneous fat distribution in the liver. Figure 8 (last column) shows an example of a blood vessel marked as a metastasis. Three other false positives are located on the edge of the liver, which sometimes has a higher intensity than the rest of the liver (e.g., Fig. 3, second column). Furthermore, motion artifacts can be the cause of a false positive object, which is the case for two false positives.

The morphological operations of Sec. 3.2 were carefully designed for the tasks described. However, post hoc analysis revealed that the liver mask dilation with a 5 × 5 structuring element did not seem to be necessary for this data set since all liver metastases were included in the original liver mask. On the other hand, the morphological closing and opening of the lesion detection results did have an impact on the results of the dual pathway method. Without these morphological operations, the median FPC would increase to eight instead of two. Nevertheless, the morphological operations removed one small lesion, from a total of 86, which was otherwise detected.

These CNN models are trained on images from a specific MR protocol from one clinic. Like other CNN models, there might be a drop in performance when the model is used on data from another scanner or another clinic. To avoid this problem, the CNN should be fine-tuned on data similar to the test data, e.g., using transfer learning.

This work could also be expanded to 3-D input images, adding more spatial information, which might improve the results. However, this would require more training data and computational power to train and test a network with more parameters.

7 Conclusions

An automatic liver segmentation with a similarity index comparable to that of the interobserver agreement is obtained from DCE MR images with an FCN. The method can accurately segment livers irrespective of the liver condition or the presence of lesions. Only in the event of lesions larger than the receptive field of the FCN are parts of the liver missed.

The proposed dual pathway metastases detection method, based on DCE MR and DW MR images, successfully detects 99.8% of the liver metastases at the cost of a median of two FPC. This will aid radiologists in locating metastases quickly.

Disclosures

The authors have no conflicts of interest to declare.

Acknowledgments

The authors thank Dr. Ashis Kumar Dhara and Professor Dr. Robin Strand from the Centre of Image Analysis of Uppsala University for discussions regarding the lesion detection.
method. This work was financially supported by the project IMPACT ("Intelligence based iMprovement of Personalized treatment And Clinical workflow support") in the framework of the EU research programme ITEA3 (Information Technology for European Advancement).

References

Mariëlle J. A. Jansen is a PhD candidate at the Image Sciences Institute at the UMC Utrecht. She received her MSc degree in biomedical engineering at the University of Twente in 2015. Her current research focuses on image processing and image analysis of focal liver lesions.

Hugo J. Kuijf is an assistant professor at the Image Sciences Institute of the UMC Utrecht. His research focuses on innovative image processing and (deep) machine learning techniques for the quantification and assessment of brain MR images. These techniques are evaluated and applied in the context of (vascular) brain pathology and dementia research, for which he collaborates in (international) multidisciplinary teams.

Maarten Niekel, MD has been a radiologist since 2017. He is currently working in the Antwerp University Hospital, after doing a fellowship at the UMC Utrecht. He worked on several projects regarding detection of colorectal metastases at the Amsterdam UMC and various projects at the Massachusetts Medical Hospital in Boston.

Wouter B. Veldhuis, PhD MD, is a radiologist holding a PhD in MR imaging and spectroscopy, who has authored over 130 papers (bit.ly/PubWV). Before coming to the UMC Utrecht, he was an oncologic imaging scholar at the MSKCC New York and a postdoc at the Stanford University Richard M. Lucas MRI Center. He is the initiator of IMAGr, the image analytics infrastructure that brings deep-learning algorithms directly into the clinical workflow in the UMC Utrecht.

Frank J. Wessels, MD, completed his radiology training in 2014, followed by a 1-year subspecialty training in abdominal radiology.
He currently holds a position as a consultant radiologist at the UMC Utrecht and the Central Military Hospital in Utrecht. His main focus is on abdominal radiology and nonvascular interventional procedures. In addition to his clinical work, he actively participates in both scientific and educational projects.

Max A. Viergever is a professor emeritus of medical imaging at Utrecht University and founder of the Image Sciences Institute of UMC Utrecht. He has (co)authored more than 750 refereed scientific articles, (co)authored/edited 18 books, and supervised more than 150 PhD theses. He was editor-in-chief of the IEEE Transactions on Medical Imaging from 2002 to 2008. He is an honorary senator of the University of Ljubljana and has received numerous career awards including the IEEE-EMBS Academic Career Achievement Award and the MICCAI Enduring Impact Award.

Josien P. W. Pluim is a professor of medical image analysis at Eindhoven University of Technology and holds a part-time appointment at the UMC Utrecht. She received her MSc degree in computer science from the University of Groningen in 1996 and her PhD from UMC Utrecht in 2001. She was the chair of SPIE Medical Imaging Image Processing from 2006 to 2009, is an associate editor of four international journals (including the Journal of Medical Imaging), and is a member of the MICCAI board of directors.