🏛 Thieme

A 3D Deep Neural Network for Liver Volumetry in 3T Contrast-Enhanced MRI

Verwendung eines 3D-neuronalen Netzwerkes zur Lebervolumenbestimmung in der kontrastmittelverstärkten 3T-MRT

Authors
Hinrich Winther[1], Christian Hundt[2], Kristina Imeen Ringe[1], Frank K. Wacker[1], Bertil Schmidt[2], Julian Jürgens[3], Michael Haimerl[4], Lukas Philipp Beyer[4], Christian Stroszczynski[4], Philipp Wiggermann[5], Niklas Verloh[4]

Affiliations
1 Department of Diagnostic and Interventional Radiology, Hannover Medical School, Hannover, Germany
2 Institute for Computer Science, Johannes Gutenberg University, Mainz, Germany
3 Department of Radiology and Nuclear Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
4 Department of Radiology, University Hospital Regensburg, Regensburg, Germany
5 Department of Radiology and Nuclear Medicine, Hospital Braunschweig, Germany

Correspondence
Dr. Niklas Verloh
Department of Radiology, University Hospital Regensburg,
Franz-Josef-Strauß-Allee 11, 93053 Regensburg, Germany
Tel.: ++49/9 41/9 44 74 01
Fax: ++49/9 41/9 44 74 02
Niklas.Verloh@klinik.uni-regensburg.de

ZUSAMMENFASSUNG

Ziel Ziel dieser Studie war es, eine vollautomatische und zuverlässige Lebervolumetrie in der kontrastverstärkten MRT basierend auf 3D-Deep-Learning-Algorithmen zu entwickeln.
Material und Methoden Datensätze von Gd-EOB-DTPA-verstärkten Leber-MR-Bildern von 100 Patienten wurden von einem in der hepatobiliären Bildgebung erfahrenen Radiologen manuell segmentiert und als Grundwahrheitssegmentierung angenommen. Die Datensätze wurden mittels einem Kreuzvalidierungsverfahren (k = 4) in Trainings- und Validierungsdatensatz eingeteilt und einem neuronalen Netzwerk zur automatischen Bildsegmentierung zugeführt. Zusätzlich wurde ein Teil der Daten (n = 9) von einem zweiten Radiologen zur Bestimmung einer Interobserver Variability segmentiert.
Ergebnisse Die manuelle Segmentierung erreichte einen Inter-Klassen-Korrelationskoeffizienten (ICC) von 0,973, einen Sørensen-Dice-Index von 95,2 ± 2,8 % und eine Überlappung von 90,9 ± 4,9 %. Das neuronale Netzwerk erreichte einen ICC von 0,98, einen Sørensen-Dice-Index von 96 ± 1,9 % und eine Überlappung von 92 ± 3,5 % sowie eine Hausdorff-Distanz von 24,9 ± 14,7 mm.
Schlussfolgerung Diese Studie präsentiert ein vollautomatisches Lebervolumetrie-Schema für MR-Bildgebung. Das neuronale Netzwerk erreichte eine kompetitive Übereinstimmung mit der Grundwahrheit bezüglich ICC, Sørensen-Dice-Index und Überlappung im Vergleich zu einer manuellen Segmentierung. Das neuronale Netzwerk erledigte die Aufgabe in nur 60 Sekunden.

Kernaussagen:
- Das vorgeschlagene neuronale Netzwerk hilft bei der genauen Segmentierung der Leber und liefert detaillierte Informationen über die patientenspezifische Anatomie und das Volumen der Leber.
- Mithilfe eines neuronalen Netzes kann eine vollautomatische Segmentierung der Leber in MRT-Scans in Sekundenschnelle durchgeführt werden.
- Ein vollautomatisches Segmentierungsschema macht die Lebersegmentierung in der MRT zu einem wertvollen Instrument für die Behandlungsplanung.

ABSTRACT

Purpose To create a fully automated, reliable, and fast segmentation tool for Gd-EOB-DTPA-enhanced MRI scans using deep learning.
Materials and Methods Datasets of Gd-EOB-DTPA-enhanced liver MR images of 100 patients were assembled. Ground truth segmentation of the hepatobiliary phase images

was performed manually. Automatic image segmentation was achieved with a deep convolutional neural network.

Results Our neural network achieves an intraclass correlation coefficient (ICC) of 0.987, a Sørensen–Dice coefficient of 96.7 ± 1.9 % (mean ± std), an overlap of 92 ± 3.5 %, and a Hausdorff distance of 24.9 ± 14.7 mm compared with two expert readers who corresponded to an ICC of 0.973, a Sørensen–Dice coefficient of 95.2 ± 2.8 %, and an overlap of 90.9 ± 4.9 %. A second human reader achieved a Sørensen–Dice coefficient of 95 % on a subset of the test set.

Conclusion Our study introduces a fully automated liver volumetry scheme for Gd-EOB-DTPA-enhanced MR imaging. The neural network achieves competitive concordance with the ground truth regarding ICC, Sørensen–Dice, and overlap compared with manual segmentation. The neural network performs the task in just 60 seconds.

Key Points:
- The proposed neural network helps to segment the liver accurately, providing detailed information about patient-specific liver anatomy and volume.
- With the help of a deep learning-based neural network, fully automatic segmentation of the liver on MRI scans can be performed in seconds.
- A fully automatic segmentation scheme makes liver segmentation on MRI a valuable tool for treatment planning.

Citation Format
- Winther H, Hundt C, Ringe KI et al. A 3D Deep Neural Network for Liver Volumetry in 3T Contrast-Enhanced MRI. Fortschr Röntgenstr 2021; 193: 305–314

Introduction

Assessment of hepatic functional reserve is crucial for the prediction of prognosis and clinical management of patients with chronic liver disease and patients undergoing liver surgery [1, 2]. Preoperative liver volumetry is essential for patient selection [3].

Liver segmentation remains challenging for computer-based approaches due to high variability in liver size and shape, the low contrast between adjacent tissues or organs, and the presence of pathologies, e. g., tumor or cirrhosis [4]. Current methods for liver segmentation are optimized for CT imaging. In most cases, segmentation is performed by a radiologist delineating a free-hand contour of the liver outline. This approach is time-consuming and might overestimate the liver volume, as the calculation includes all structures within the contours of the segmentation, e. g., vessels or liver tumors. Semi- and fully automated systems have been introduced to support liver volume estimation in CT scans [5–12], although liver segmentation on magnetic resonance imaging (MRI) remains mostly unexplored. However, MR imaging has been increasingly used for liver resection and transplantation planning, leading to demands for developing computerized MRI liver volumetry approaches [13].

Huynh et al. [14] proposed a fully automated scheme for liver segmentation based on watershed transformation in combination with active contouring on scanners. The authors determined an intraclass correlation coefficient of 0.94 with an average execution time of 8.4 min per case [14]. In contrast, we analyzed contrast-enhanced images with Gd-EOB-DTPA, a liver-specific contrast agent, while Huynh et al. used a non-liver-specific contrast agent.

Gd-EOB-DTPA is a liver-specific contrast agent. Its biochemical properties allow general assessment of tissue perfusion in the vascular phase in addition to its specific accumulation in hepatocytes in the late phase (hepatobiliary phase) after approximately 15–20 minutes [15–22]. Gadolinium uptake shortens the spin-lattice relaxation time in the corresponding tissue, which leads to increased signal intensity (SI) on T1-weighted images. Typically, segmentation quality is heavily reliant on the contrast of the SI distribution, and thus, using Gd-EOB-DTPA-enhanced MRI is expected to boost segmentation performance [16, 18, 22, 23].

This study aims to create a fully automated, reliable, and fast segmentation tool for Gd-EOB-DTPA-enhanced MRI scans acquired using 3 T scanners.

Materials and Methods

Patients

The Institutional Review Board approved this retrospective study and written informed consent was obtained from all patients before undergoing MRI examination. 100 Gd-EOB-DTPA-enhanced MRI liver scans were selected randomly from the clinical routine. The patients underwent liver MRI as a complementary examination due to the presence of unknown liver lesions seen in computed tomography (CT) or ultrasound or as a control examination after treatment of a focal malignant liver lesion. None of the recruited patients had any contraindications for MRI examination (e. g., claustrophobia and incompatible metallic implants), contraindications for the administration of Gd-EOB-DTPA (e. g., renal failure) or known previous reactions to liver-specific MRI contrast agents.

Imaging

Imaging was performed using a clinical whole-body system (Magnetom Skyra, Siemens Healthineers) in combination with two body-spine array coil elements (an 18-channel body matrix coil and a 32-channel spine matrix coil) for signal reception.

The following parameters were used for image acquisition: a T1-weighted volume-interpolated breath-hold examination (VIBE) sequence with fat suppression (repetition time (TR): 3.09 ms; echo time (TE): 1.16 ms; flip angle: 9; parallel imaging factor: 2; slices: 64; reconstructed voxel size: $1.3 \times 1.3 \times 3.0$ mm^3; measured voxel size: $1.7 \times 1.3 \times 4.5$ mm^3; acquisition time: 14 s),
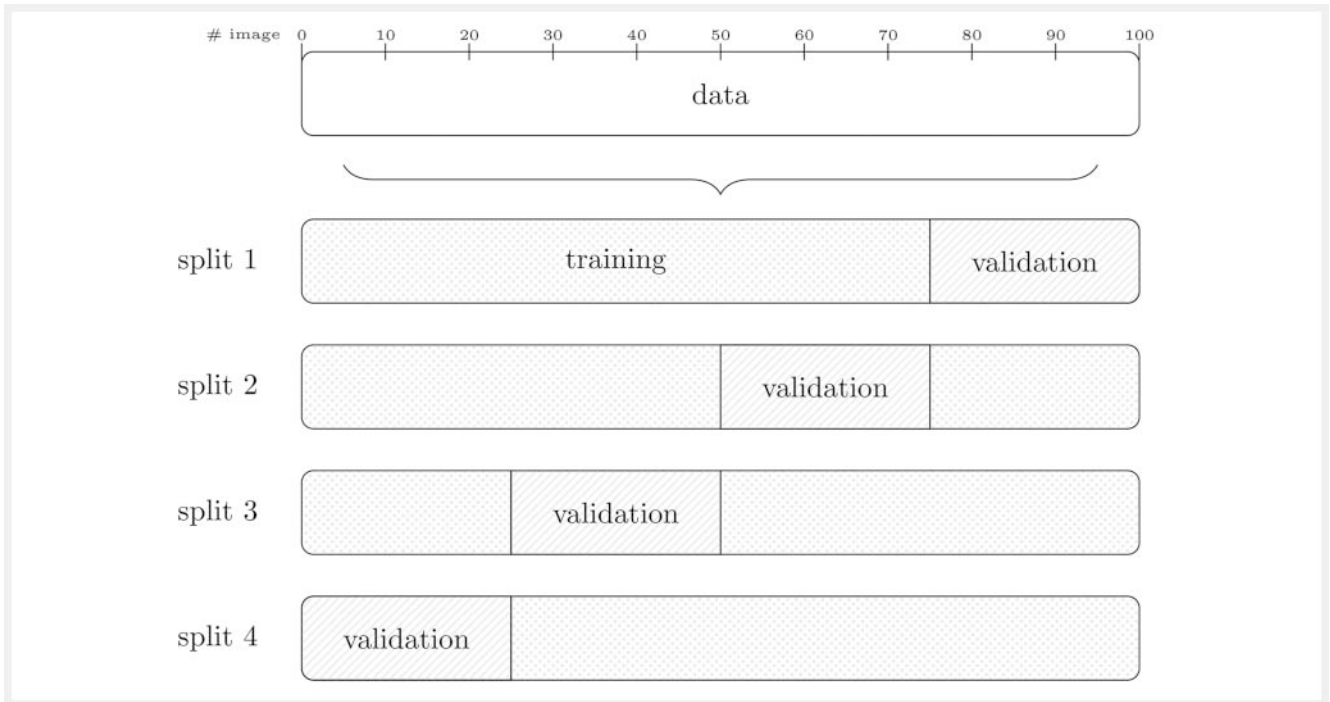
▶ **Fig. 1** Four-fold cross-validation. The 100 cases are divided into four equally sized portions. Subsequently, three of the four portions are used for training, and one for validation. In a four-fold cross-validation scheme, there are four possible ways to split the data into training and validation sets. Each validation set is composed of mutually exclusive cases. Quality measures are determined for each of the four possible splits, each containing 75 training and 25 validation cases.

▶ **Abb. 1** Vierfache Kreuzvalidierung. Die 100 Fälle wurden in 4 gleich große Teile aufgeteilt. Anschließend wurden jeweils 3 der 4 Teile für das Training und einer für die Validierung verwendet. In einem 4-fachen Kreuzvalidierungsschema gibt es 4 Möglichkeiten, die Daten in Trainings- und Validierungssätze aufzuteilen. Jeder Validierungssatz besteht aus sich gegenseitig ausschließenden Fällen. Für jede der 4 möglichen Aufteilungen mit jeweils 75 Trainings- und 25 Validierungsfällen wurde die Auswertung durchgeführt.
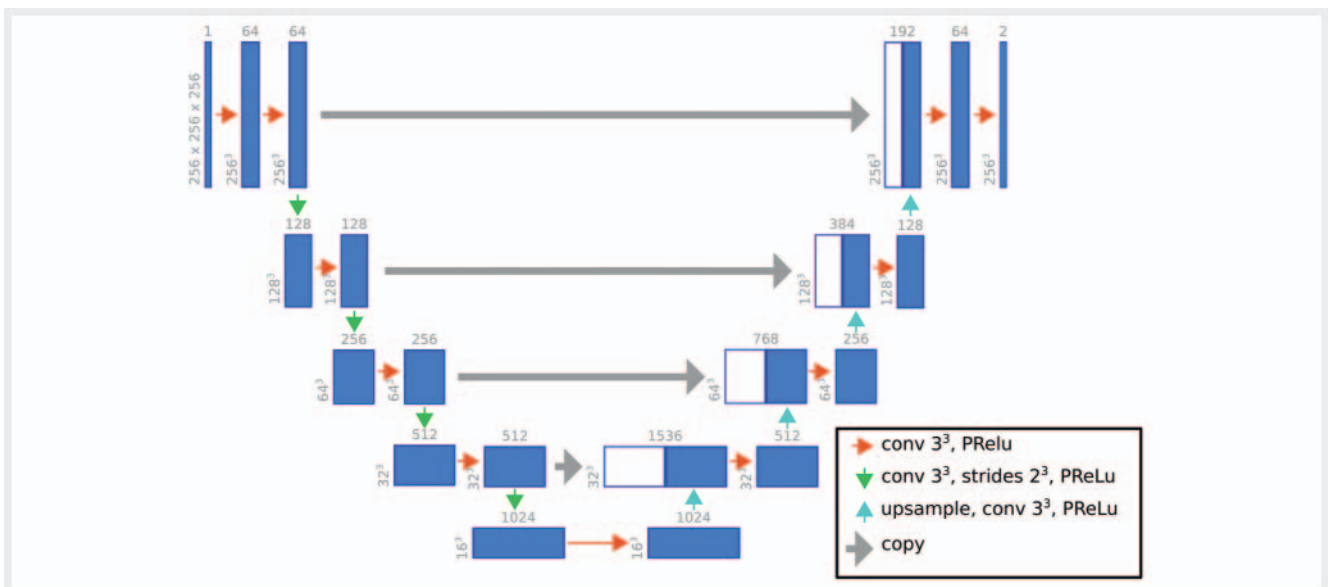
▶ **Fig. 2** Topology of the neural network. The figure represents the topology of the neural network. Each blue box represents a multichannel feature map. The size and number of channels are denoted in the upper and left corners. Color-coded arrows illustrate the different operations.

▶ **Abb. 2** Topologie des neuronalen Netzes. Die Abbildung stellt die Topologie des neuronalen Netzes dar. Jede blaue Box stellt eine Mehrkanal-Ausgabe dar. Die Größe und Anzahl der Kanäle sind in der oberen und linken Ecke angegeben. Farbkodierte Pfeile veranschaulichen die verschiedenen Operationen.

covering the entire liver, acquired during one breath-hold in the hepatobiliary phase (20 minutes after contrast injection).

All patients received a Gd-EOB-DTPA (Primovist, Eovist; Bayer Schering Pharma AG, Berlin, Germany) dose that was adapted to their respective body weights (0.025 mmol/kg). The hepatocyte-specific contrast agent Gd-EOB-DTPA was administered via bolus injection with a flow rate of 1 ml/s and flushed with 20 ml of a 0.9 % saline solution.

Image Analysis

The ground truth segmentation of the corresponding MR images in the hepatobiliary phase was performed by a senior physician (5 years of experience) for all 100 cases according to the current gold standard of manual liver segmentation. Liver lesions were excluded from volumetry. This process was supported by a semi-automated region growing algorithm featuring manual edge correction as implemented by the open-source software Osirix (version 7.5) [24].

Furthermore, 9 of the training images were segmented by a second reader (5 years of experience) to estimate the inter-expert/inter-program agreement in terms of the Dice index, overlap, and intraclass correlation coefficient (ICC) of the liver volume. This reflects the typical use case of liver segmentation by different experts using different software suites. Furthermore, this avoids confounding issues regarding semiautomatic segmentation algorithms, as the second segmentation was performed using a different method, based on active contour (also known as "SNAKE") with manual edge correction, as implemented by itk-SNAP (version 3.5) [25].

Dataset

The segmentation quality was determined using a four-fold cross-validation protocol, as depicted in ▶ **Fig. 1**. This method offers the advantage of using the complete dataset for validation. This allows for a smaller dataset in total and, in turn, less manual and time-consuming image segmentation.

Network Topology (3 D)

The neural network is a derivative of 3 D U-Net [26]. The architecture consists of a downsampling cascade of layers for feature extraction and subsequent symmetric upsampling cascade that enables precise localization. In contrast to the original implementation, we apply the lessons learned from our previous study [27] and substitute the rectified linear unit (ReLu) activation function with a parametric rectified linear unit (PReLu). Grayscale images with a size of 96 × 96 × 96 are fed to the input layer and subsequently downsampled using 3 × 3 × 3 convolutions with striding of 2 × 2 × 2. Upsampling is performed by 3 × 3 × 3 convolution after bilinear interpolation to double resolution. The input layer operates on a resolution of 1.5 × 1.5 × 1.5 mm. An ensemble strategy with an overlap of 0.5 is used to infer the total volume. A Gaussian distribution is used to emphasize the central field of view portions of the prediction in regard to the outer edges. The neural network topology of this study is depicted in ▶ **Fig. 2**.
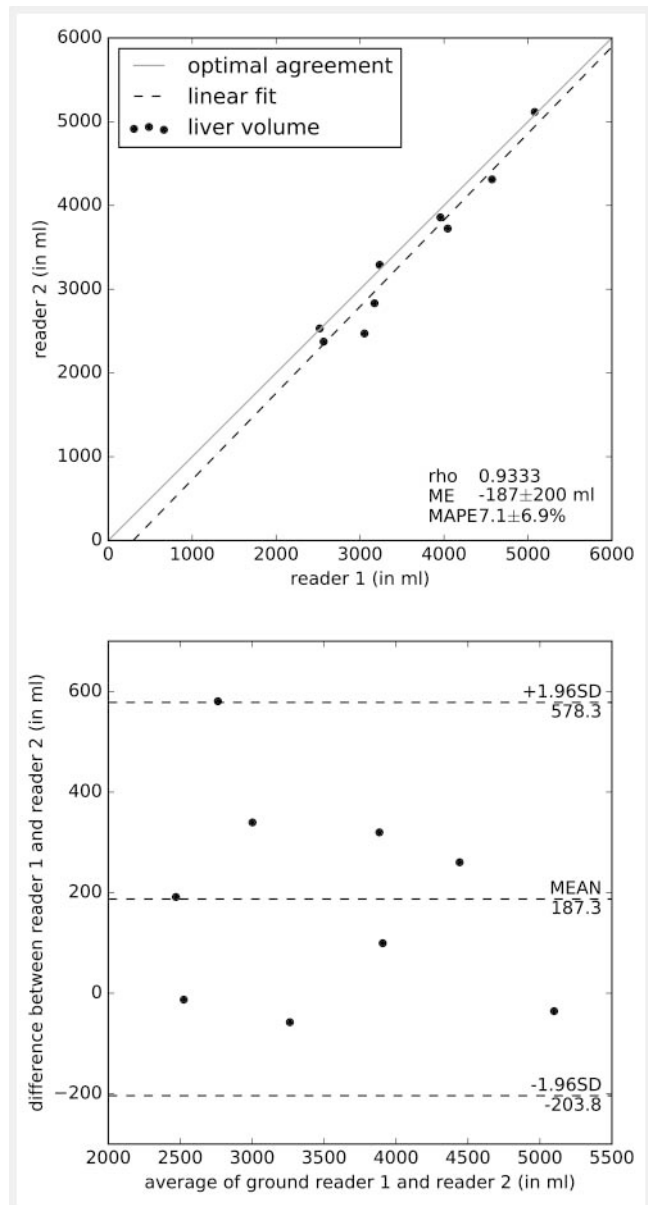
▶ **Fig. 3** Visualization of the inter-reader/inter-program variance. The upper panel shows a basic correlation plot where the abscissa (x-axis) corresponds to the liver volume segmented by the first reader, and the ordinate (y-axis) corresponds to the liver volume segmented by the second reader. Spearman's rank correlation coefficient (rho), mean error (ME) and mean absolute percentage error (MAPE) were calculated for the inter-reader/inter-program analysis. The lower panel shows the associated Bland-Altman plot of the liver volumes.

▶ **Abb. 3** Visualisierung der Interreader/Interprogramm-Varianz. Das obere Panel zeigt einen Korrelationsplot, bei dem die Abszisse (x-Achse) dem vom ersten Auswerter segmentierten Lebervolumen und die Ordinate (y-Achse) dem vom zweiten Auswerter segmentierten Lebervolumen entspricht. Spearmans Korrelationskoeffizient (rho), mittlerer Fehler (ME) und mittlerer absoluter prozentualer Fehler (MAPE) wurden für die Interreader/Interprogramm-Analyse berechnet. Das untere Panel zeigt den dazugehörigen Bland-Altman-Plot der Lebervolumina.

▶ **Table 1** Comparative studies. Comparison of our results to previous work. The table depicts overlap metrics, such as overlap, Sørensen–Dice coefficient (Dice), and Hausdorff distance (dH), and the intraclass correlation coefficient (ICC). The best results in the fully automated bracket are denoted in bold.

▶ **Tab. 1** Vergleichende Studien. Vergleich unserer Ergebnisse mit früheren Arbeiten. Die Tabelle zeigt Überlappungsmetriken wie Überlappung, Sørensen–Dice Koeffizient (Dice) und Hausdorff-Distanz (dH) sowie den Intraklassen-Korrelationskoeffizienten (ICC). Die besten Ergebnisse in der vollautomatischen Gruppe sind fettgedruckt.

method	n	overlap (%)	dice (%)	ICC	dH
manual					
▪ inter-reader	9	90.9 ± 4.9	95.2 ± 2.8	0.973	31.6 ± 5.9
Semiautomated					
▪ Chartrand et al. [40]	21	89.3 ± 2.9	n/a	n/a	n/a
▪ Chartrand et al. [36]	20	92.4 ± 1.4	n/a	n/a	n/a
fully automated					
▪ Chen et al. [46]	n/a	n/a	≈ 80	n/a	n/a
▪ Ruskó and Bekes [47]	8	88.8 ± 4.1	94.1 ± 2.3	n/a	n/a
▪ Huynh et al. [14]	23	n/a	93.6 ± 1.7	0.98	12.8 ± 2.24*
▪ López-Mir et al. [48]	17	n/a	95 ± 1.5	n/a	33.6 ± 6.1
▪ Bereciartua et al. [49]	18	n/a	90.2 ± 8.6	n/a	**22.7 ± 12.0**
▪ Yan et al. [41]	14	n/a	86 ± 5	n/a	≈ 32**
▪ Huynh et al. [15]	27	n/a	91.1 ± 1.9	0.94	n/a
fully automated, based on the proposed					
▪ 3 D neural network	**100**	**92.3 ± 3.5**	**96.0 ± 1.9**	**0.987**	24.9 ± 14.7

* modified Hausdorff distance, not directly comparable to the otherwise reported original Hausdorff distance.
  modifizierte Hausdorff-Distanz, nicht direkt vergleichbar mit der sonst berichteten ursprünglichen Hausdorff-Distanz.
** Yan et al. [41] did not explicitly report the Hausdorff distance; however, they depict it in a figure. The aforementioned value is the best approximation by the authors.
  Yan et al. [41] nennen nicht explizit die Hausdorff-Distanz; sie stellen diese jedoch in einer Abbildung dar. Der oben genannte Wert ist eine Annäherung.

Network Training

The objective of the neural network during training is to minimize the weak binary cross-entropy between predicted and ground truth segmentation. The weights are initialized with random values and sampled from a uniform distribution without scaling variance (uniform scaling), as proposed by [28]. The adaptive moment estimation (ADAM) algorithm is used for stochastic optimization [29]. The initial learning rate of 10^{-3} was gradually reduced to 10^{-6}. Extensive nonlinear image augmentation was performed during training, as described in [27].

In our experiments, we use a six-core Intel i7-5930K CPU @ 3.50 GHz with 32 GB of RAM and a Maxwell-based NVIDIA GeForce GTX TITAN X (GM200) accelerator board with 12 GB video RAM.

Statistical Analysis

Statistical analysis was performed with Python 2.7 and R 3.2 [30]. Statistical measures are reported in the format μ ± σ, where μ is the ensemble mean, and σ is the standard deviation (SD) in percent. The ICC, Sørensen–Dice coefficient (Dice), and overlap were calculated between the two expert readers and between human and machine. Overlap $|A{\cap}B|/|AB|$ is defined as the quotient of the intersection volume of both segmentations A, B and the volume of their unification. Note that A corresponds to the predicted segmentation, and B corresponds to the ground truth segmentation in our experiments.

Results

Interreader/Interprogram Variance

The inter-reader/inter-program correlation is depicted in ▶ **Fig. 3**. The two expert readers account for an ICC of 0.973, a Dice index of 95.2 ± 2.8 %, an overlap of 90.9 ± 4.9 %, with a mean absolute volume difference of 187 ± 200 ml (6.4 ± 7.5 % relative volume) and a mean absolute percentage error (MAPE) of 7.1 ± 6.9 %.

Automated Liver Segmentation

The neural network achieves an ICC of 0.987, a Dice coefficient of 96.0 ± 1.9 %, and an overlap of 90.0 ± 3.6 %, as depicted in ▶ **Table 1**. Comparing the volumes to the manual segmentation, the proposed method achieved a mean absolute difference of
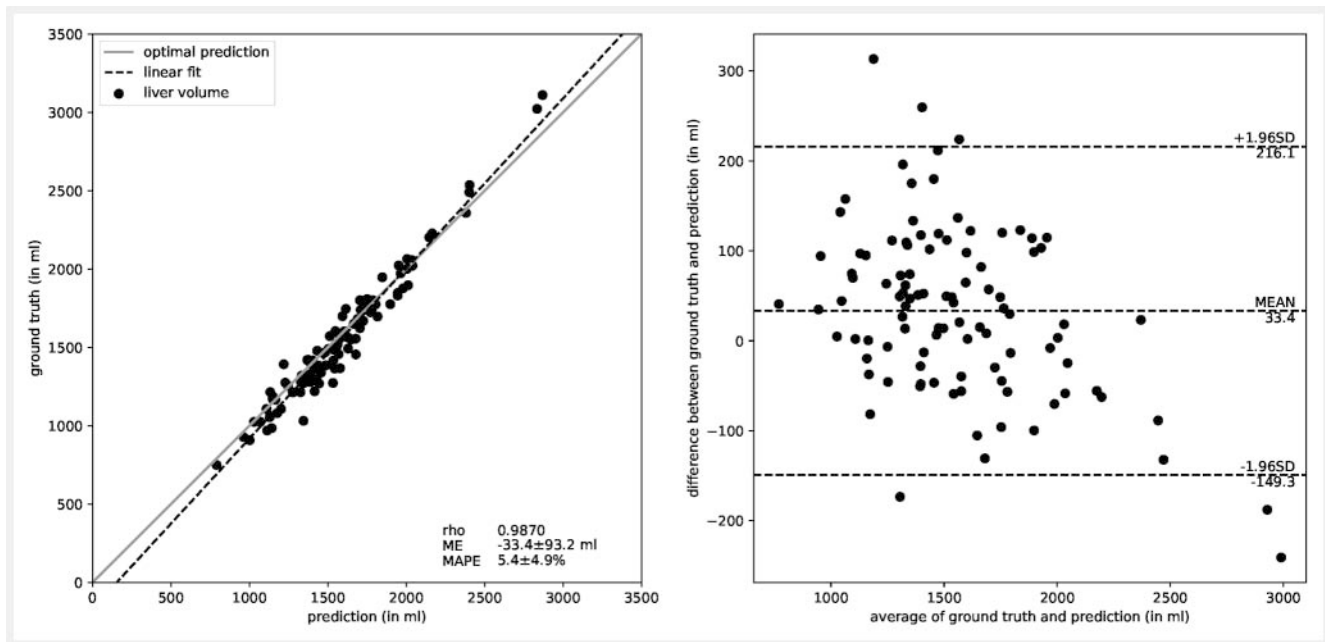
▶ **Fig. 4** Visualization of agreement between the ground truth and predicted volumes. The left panel shows a basic correlation plot where the abscissa (x-axis) corresponds to predicted volumes and the ordinate (y-axis) to ground truth volumes. The optimal linear regression (dashed line) is almost identical to the identity map (solid main diagonal). Spearman's rank correlation coefficient (rho), mean error (ME), and mean absolute percentage error (MAPE) were calculated over all 4 × 25 predicted volumes in the four splits of the four-fold cross-validation. The right panel shows the associated Bland-Altman plot of the liver volumes.

▶ **Abb. 4** Visualisierung der Übereinstimmung zwischen der Grundwahrheit und den vorhergesagten Volumina. Die linke Abbildung zeigt ein Korrelationsdiagramm, bei dem die Abszisse (x-Achse) den vorhergesagten Volumina und die Ordinate (y-Achse) der Grundwahrheit-Volumina entspricht. Die optimale lineare Regression (gestrichelte Linie) ist fast identisch mit der Identitätskarte (durchgezogene Hauptdiagonale). Spearman Korrelationskoeffizient (rho), mittlerer Fehler (ME) und mittlerer absoluter prozentualer Fehler (MAPE) wurden über alle 4 × 25 vorhergesagten Volumina in den 4 Aufteilungen der 4-fachen Kreuzvalidierung berechnet. Die rechte Tafel zeigt den zugehörigen Bland-Altman-Plot der Lebervolumina.

33.4 ± 93.2 ml (2.2 ± 6.6 % relative volume) with a MAPE of 5.4 ± 4.9 % (▶ **Fig. 4**). No relevant correlation (Spearman rho: 0.3) was found between the quality of segmentation (measured with overlap) and the mean signal intensity of the liver volume (▶ **Fig. 5**).

Segmentation Time

The neural network has a fixed input size of 96^3. A typical MR image has a matrix shape of $320 \times 320 \times 64$ with a spatial resolution of $1.25 \times 1.25 \times 3$ mm. After normalizing the image to a spatial resolution of 1.5^3 mm, the actual matrix shape is $267 \times 267 \times 128$ for the specific example. This shape is incompatible with the fixed size of the input layer. To accommodate this issue, we employ an ensemble strategy by inferring 96^3 blocks with an overlap of 50 %. A Gaussian distribution is used to emphasize the central field of view portions of the prediction in regard to the outer edges. Therefore, the fully automated prediction step corresponds to multiple feedforward passes of the neural network performed on the GPU. On average, it takes approximately 60 seconds to predict the segmentation of a whole MRI scan of the aforementioned dimensions. In contrast, manual segmentation, performed by a domain expert, accounts for approximately 10 ± 2 minutes per case.

Discussion

3 D segmentation in medical image analysis is a time-consuming and tedious task. Liver MRI segmentation remains challenging due to the lack of well-established, fully automated frameworks. However, this is a mandatory preliminary step before performing liver resection and transplantation and is executed on a daily basis. The design and implementation of robust and efficient segmentation algorithms are, therefore, of high importance to the clinical routine. Many approaches rely on basic image processing techniques, such as thresholding of intensity values, histograms, and morphological operations in combination with complementary methods, such as atlas-guided algorithms, region growing approaches, deformable models, and classification-based methods from the field of machine learning [31–34].

Established segmentation procedures for CT imaging have been translated or adjusted to MR imaging. However, due to major differences in the image morphology, these approaches may not be directly applicable to MR imaging with comparable segmentation performance [35].

Chartrand et al. [35] proposed a semi-automated segmentation method for CT scans and MR images using Laplacian mesh optimization. This approach achieves an overlap of 92.4 ± 1.4 %
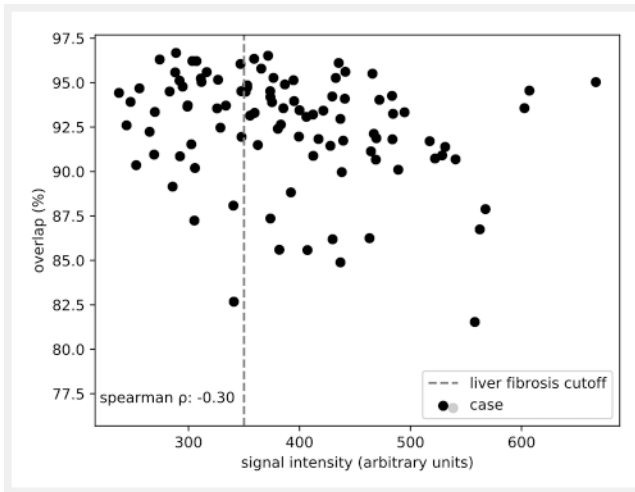
► **Fig. 5** Visualization of the agreement based on the level of signal intensity. The figure shows a scatterplot in which the abscissa (x-axis) corresponds to the mean signal intensity of the liver parenchyma and the ordinate (y-axis) to the agreement, measured by the overlap in %, between the ground truth and the predicted segmentation. The Spearman correlation coefficient (rho) was calculated between the overlap and the mean signal intensity of the liver volume.

► **Abb. 5** Visualisierung der Übereinstimmung in Abhängigkeit der Signalintensität. Die Abbildung zeigt ein Streudiagramm, bei dem die Abszisse (x-Achse) der mittleren Signalintensität des Leberparenchyms und die Ordinate (y-Achse) der Übereinstimmung, gemessen anhand der Überlappung in % zwischen der Grundwahrheit und der vorhergesagten Segmentierungen, entspricht. Der Spearman Korrelationskoeffizient (rho) wurde zwischen der Überlappung und der mittleren Signalintensität des Lebervolumens berechnet.

for MR imaging. However, it must be stressed that the semi-automated method requires manual interaction.

Cheng et al. [36] proposed a fully automated process based on level-sets using a prior shape model, whereas Chan and Vese [37] combined active shape recognition with region growing. These approaches are prone to error for strong noise or artifacts inside the liver parenchyma and low contrast due to missing edge recognition [12].

Gloger et al. [38] discussed a fully automated segmentation procedure combining region growing and a threshold-based technique. Chartrand et al. [39] described a semi-automated method that combines minimal path surface segmentation with model deformation. Huynh et al. [14] proposed a fully automated scheme using watershed segmentation coupled with active contouring. These protocols present promising results, with a Dice coefficient of up to 95 %. However, most of these protocols are time-consuming, with a runtime of approximately 5–8 minutes per case [14, 38, 39].

Deep convolutional neural networks have outperformed state of the art techniques in most visual recognition tasks [40]. This technique was recently applied to CT liver segmentation by Hu et al. [41] and Lu et al. [42] with good results in terms of volume overlap metrics (Dice index ≈ 97 % and overlap error ≈ 6 %). However, deep learning has yet to be applied to MRI liver segmentation.

Our approach uses a 3D evolution of the deep neural network v-net [27]. It achieves a liver segmentation performance comparable to human experts with an overlap of 92 %/91 % (neural network/human experts), a Dice coefficient of 96 %/95 %, and an ICC of 0.99/0.97. However, the total volume difference is lower for the proposed method (2 %) than for human experts (6 %), while the segmentation process is significantly faster with approximately 60 seconds per case, compared with an average of 10 minutes per case when performed manually. The results suggest comparable or higher segmentation performance compared with the aforementioned semi- and fully automated approaches.

We use semiautomatic, algorithm-based, manual segmentation as the ground truth. In a subsequent visual comparison of prediction and ground truth segmentation, we found subjectively less accurate segmentation quality for manual segmentation in some of the upper and lower regions of the liver, typically the extremes of segments VI and VIII, and in proximity of intrahepatic lesions or vessels (► **Fig. 6**). However, this finding cannot be objectively measured due to the lack of flawless ground truth segmentation. One explanation for the imprecise segmentation quality may be the time factor. The segmentations were performed as part of the clinical routine, and thus, time was limited. One of the most remarkable qualities of humans is the active and dynamic way in which we process information [43]. However, it can also influence and distort how we perceive visual information and how we make decisions [44]. Human work thus tends to be imprecise, due to the natural tendencies of distraction, attention deficit, fatigue, disruption of a decision, analysis of incomplete information, and distortion.

Liver segmentation was performed on T1-weighted images using the hepatobiliary phase, expecting a boost in performance due to the specific contrast uptake of hepatocytes. In patients with normal liver function, this leads to a significant increase in the liver parenchyma's signal intensity [45, 46]. Within these patients, segmentation of the liver, parenchyma is not a challenge. In patients with liver fibrosis or even cirrhosis, the liver parenchyma's signal intensity is decreased depending on the degree of liver fibrosis [47]. Within the analyzed dataset of 100 randomly selected cases, patients with low parenchyma contrast, in terms of liver fibrosis, were included. In a subgroup investigation, we analyzed the quality of segmentation (measured with the overlap) and the mean signal intensity of the liver volume. No relevant correlation was found between the signal intensity of the liver parenchyma and the overlap (Spearman rho: 0.3), indicating robust segmentation of the liver parenchyma even in reduced uptake of Gd-EOB-DTPA in hepatocytes. In cases with reduced uptake of Gd-EOB-DTPA, manual segmentation took considerably longer, whereas the network was able to segment these scans quickly.

Benign liver lesions are often identified initially on a (contrast-enhanced) abdominal ultrasound scan or CT scan. Supplementary MRI examinations, in particular with liver-specific contrast medium, are mainly used in the case of contradictory statements. This leads to an underrepresentation of benign liver lesions in this patient group. In the case of focal nodular hyperplasia (FNH), there was only one case within the dataset. Due to the four-fold cross-validation, the model that performed the prediction for the one FNH case had not a single training example for an FNH lesion.
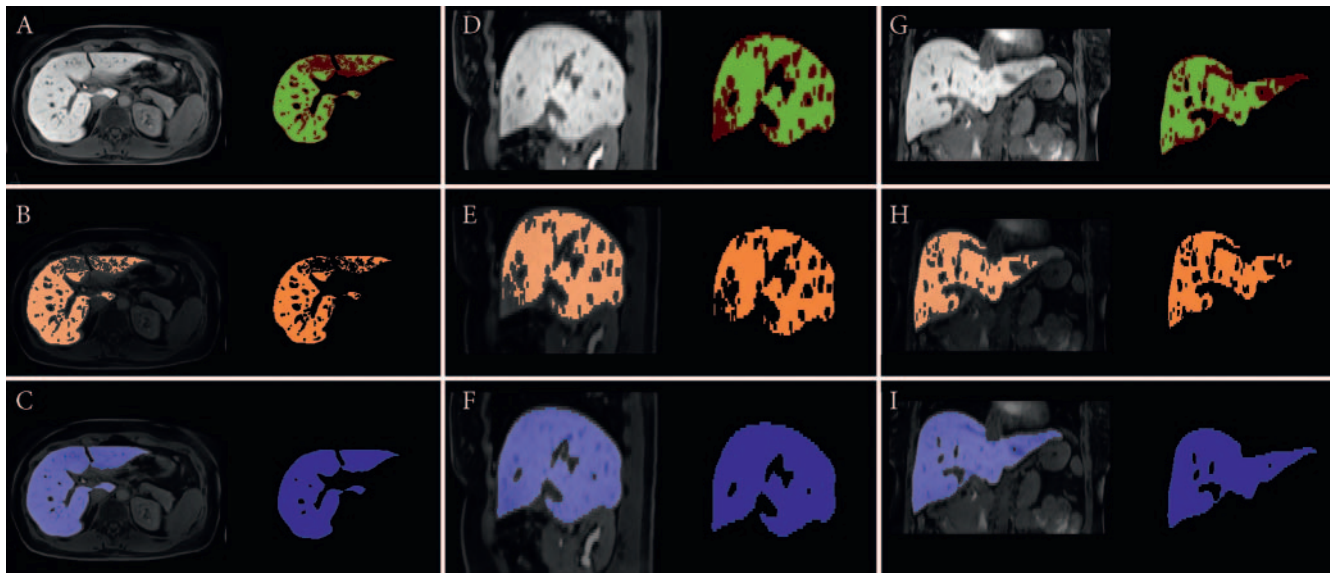
▶ **Fig. 6** Comparison between the neural network and the gold standard. Panel **a, d**, and **g** show the MRI scans of a patient with imperfect manual segmentation (**b, e, h**) for axial (**a, b, c**), sagittal (**d, e, f**), and coronal (**g, h, i**) orientations, illustrating the agreement between manual and fully automated segmentation. Green pixels correspond to areas of agreement, whereas red pixels denote areas of disagreement. Panels **c, f**, and **i** show examples of segmentation produced by the fully automated liver segmentation framework. In a visual comparison, one can clearly see the difference. Segmentation by the neural network produces a smoother result than manual segmentation.

▶ **Abb. 6** Vergleich zwischen den Ergebnissen des neuronalen Netzwerks und der Goldstandard-manuellen Segmentierung. Dargestellt sind die MTR-Aufnahmen (**a, d** und **g**) eines Patienten mit teilweiser fehlerhafter manueller Segmentierung. Abgebildet sind die axiale (**a, b, c**), die sagittale (**d, e, f**) und die koronare Ausrichtung (**g, h, i**). Zusätzlich wird die Übereinstimmung zwischen manueller und vollautomatischer Segmentierung veranschaulicht. Grüne Pixel entsprechen Übereinstimmungen, rote Pixel Abweichungen. Die Abbildungen **c, f** und **i** zeigen eine Beispielsegmentierung, die durch die vollautomatische Lebersegmentierung erzeugt wurde. In einem visuellen Vergleich kann man den Unterschied deutlich erkennen: Die Segmentierung durch das neuronale Netz erzeugt eine glattere Oberflächenstruktur als die manuelle Segmentierung.

Most likely, due to FNH lesions having functioning hepatocytes with prolonged storage of Gd-EOB-DTPA, the model misclassified the lesion as liver parenchyma. Therefore, this study does not allow any conclusions to be drawn in regard to FNH lesions.

Another limitation of this and all the aforementioned studies is the poor comparability of the results. All MRI-based studies have been reported on a different set of training and validation cases. Therefore, although our results indicate better performance than prior state-of-the-art techniques, this benefit cannot be verified without testing the validation sets of the other studies. For this reason, we are publishing our full training and test set with the original MRI scans and the corresponding manual segmentation to enable further studies to perform a direct comparison.

Furthermore, it is known that neural networks tend to overfit the training data. Therefore, we plan to further evaluate these results in a multicenter setup in a follow-up study.

Conclusion

Our proposed deep learning-based method enables robust, fast, and fully automated liver segmentation using MR scans and reliable volume estimation with an interclass correlation coefficient of 0.987. Segmentation of the liver parenchyma can be performed in just seconds, making liver segmentation on MRI a valu-

able tool for treatment planning, especially for patients undergoing liver surgery.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Ribero D, Amisano M, Bertuzzo F et al. Measured versus estimated total liver volume to preoperatively assess the adequacy of the future liver remnant: which method should we use? Ann Surg 2013; 258: 801–806; discussion 806–807

[2] Yigitler C, Farges O, Kianmanesh R et al. The small remnant liver after major liver resection: how common and how relevant? Liver transplantation: official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society 2003; 9: S18–S25

[3] Guglielmi A, Ruzzenente A, Conci S et al. How much remnant is enough in liver resection? Digestive surgery 2012; 29: 6–17

[4] Heimann T, van Ginneken B, Styner MA et al. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE transactions on medical imaging 2009; 28: 1251–1265

[5] Li G, Chen X, Shi F et al. Automatic Liver Segmentation Based on Shape Constraints and Deformable Graph Cut in CT Images. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society 2015; 24: 5315–5329

[6] Suzuki K, Epstein ML, Kohlbrenner R et al. Quantitative radiology: automated CT liver volumetry compared with interactive volumetry and manual volumetry. Am J Roentgenol American journal of roentgenology 2011; 197: W706–W712

[7] Kumar SS, Moni RS, Rajeesh J. Automatic liver and lesion segmentation: a primary step in diagnosis of liver diseases. Signal Image Video Process 2013; 7: 163–172

[8] Goryawala M, Gulec S, Bhatt R et al. A low-interaction automatic 3D liver segmentation method using computed tomography for selective internal radiation therapy. Biomed Res Int 2014; 2014: 198015

[9] Yuqian Z, Yunlong Z, Xiaofang W et al. Fuzzy C-means clustering-based multilayer perceptron neural network for liver CT images automatic segmentation. In, 2010 Chinese control and decision conference Xuzhou, China: IEEE. 2010: 3423–3427

[10] Peng J, Hu P, Lu F et al. 3D liver segmentation using multiple region appearances and graph cuts. Med Phys 2015; 42: 6840–6852

[11] Priyadarsini S, Selvathi D. Survey on segmentation of liver from CT images. In, 2012 IEEE International conference on advanced communication control and computing technologies (ICACCCT) Ramanathapuram, India: IEEE. 2012: 234–238

[12] Gotra A, Sivakumaran L, Chartrand G et al. Liver segmentation: indications, techniques and future directions. Insights into imaging 2017; 8: 377–392

[13] Huynh HT, Karademir I, Oto A et al. Computerized liver volumetry on MRI by using 3D geodesic active contour segmentation. Am J Roentgenol American journal of roentgenology 2014; 202: 152–159

[14] Huynh HT, Le-Trong N, Bao PT et al. Fully automated MR liver volumetry using watershed segmentation coupled with active contouring. Int J Comput Assist Radiol Surg 2017; 12: 235–243

[15] Nassif A, Jia J, Keiser M et al. Visualization of hepatic uptake transporter function in healthy subjects by using gadoxetic acid-enhanced MR imaging. Radiology 2012; 264: 741–750

[16] Kobayashi S, Matsui O, Gabata T et al. Relationship between signal intensity on hepatobiliary phase of gadolinium ethoxybenzyl diethylenetriaminepentaacetic acid (Gd-EOB-DTPA)-enhanced MR imaging and prognosis of borderline lesions of hepatocellular carcinoma. European journal of radiology 2012; 81: 3002–3009

[17] Van Beers BE, Pastor CM, Hussain HK. Primovist, Eovist: what to expect? J Hepatol 2012; 57: 421–429

[18] Bluemke DA, Sahani D, Amendola M et al. Efficacy and safety of MR imaging with liver-specific contrast agent: U.S. multicenter phase III study. Radiology 2005; 237: 89–98

[19] van Montfoort JE, Stieger B, Meijer DK et al. Hepatic uptake of the magnetic resonance imaging contrast agent gadoxetate by the organic anion transporting polypeptide Oatp1. The Journal of pharmacology and experimental therapeutics 1999; 290: 153–157

[20] Weinmann HJ, Bauer H, Frenzel T et al. Mechanism of hepatic uptake of gadoxetate disodium. Academic radiology 1996; 3: S232–S234

[21] Schuhmann-Giampieri G, Schmitt-Willich H, Frenzel T et al. Biliary excretion and pharmacokinetics of a gadolinium chelate used as a liver-specific contrast agent for magnetic resonance imaging in the rat. Journal of pharmaceutical sciences 1993; 82: 799–803

[22] Weinmann HJ, Schuhmann-Giampieri G, Schmitt-Willich H et al. A new lipophilic gadolinium chelate as a tissue-specific contrast medium for MRI. Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine/Society of Magnetic Resonance in Medicine 1991; 22: 233–237; discussion 242

[23] Clement O, Muhler A, Vexler V et al. Gadolinium-ethoxybenzyl-DTPA, a new liver-specific magnetic resonance contrast agent. Kinetic and enhancement patterns in normal and cholestatic rats. Investigative radiology 1992; 27: 612–619

[24] Rosset A, Spadola L, Ratib O. OsiriX: an open-source software for navigating in multidimensional DICOM images. J Digit Imaging 2004; 17: 205–216

[25] Yushkevich PA, Piven J, Hazlett HC et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage 2006; 31: 1116–1128

[26] Çiçek Ö, Abdulkadir A, Lienkamp SS et al. 3D U-net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, et al., eds; Medical image computing and computer-assisted intervention – MICCAI 2016. Cham: Springer International Publishing; 2016: 424–432

[27] Winther HB, Hundt C, Schmidt B et al. v-net: deep learning for generalized biventricular mass and function parameters using multicenter cardiac MRI data. JACC Cardiovasc Imaging 2018; 11: 1036–1038

[28] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Yee Whye T, Mike T, eds; Proceedings of the thirteenth International conference on artificial intelligence and statistics. PMLR; 2010: 249–256

[29] Kingma DP, Ba J. Adam: a method for stochastic optimization. ArXiv e-prints. 2014

[30] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2013

[31] Freiman M, Eliassaf O, Taieb Y et al. An iterative Bayesian approach for nearly automatic liver segmentation: algorithm and validation. Int J Comput Assist Radiol Surg 2008; 3: 439

[32] Gao L, Heath DG, Kuszyk BS et al. Automatic liver segmentation technique for three-dimensional visualization of CT data. Radiology 1996; 201: 359–364

[33] Seo KS. Improved fully automatic liver segmentation using histogram tail threshold algorithms. In: Sunderam VS, van Albada GD, Sloot PMA, et al., eds, Computational science – ICCS 2005. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005: 822–825

[34] Hermoye L, Laamari-Azjal I, Cao Z et al. Liver segmentation in living liver transplant donors: comparison of semiautomatic and manual methods. Radiology 2005; 234: 171–178

[35] Chartrand G, Cresson T, Chav R et al. Liver segmentation on CT and MR using laplacian mesh optimization. IEEE Trans Biomed Eng 2017; 64: 2110–2121

[36] Cheng K, Gu L, Xu J. A novel shape prior based level set method for liver segmentation from MR images. In, 2008 International conference on information technology and applications in biomedicine Shenzhen, China: IEEE. 2008: 144–147

[37] Chan TF, Vese LA. Active contours without edges. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society 2001; 10: 266–277

[38] Gloger O, Kühn J, Stanski A et al. A fully automatic three-step liver segmentation method on LDA-based probability maps for multiple contrast MR images. Magnetic resonance imaging 2010; 28: 882–897

[39] Chartrand G, Belanger C, Gotra A et al. Accuracy of a semi-automated liver segmentation method using CT scan. In, European Congress of Radiology Vienna, Austria: ESR. 2013: 1–20

[40] Yan Z, Zhang S, Tan C et al. Atlas-based liver segmentation and hepatic fat-fraction assessment for clinical trials. Comput Med Imaging Graph 2015; 41: 80–92

[41] Hu P, Wu F, Peng J et al. Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. Phys Med Biol 2016; 61: 8676–8698

[42] Lu F, Wu F, Hu P et al. Automatic 3D liver location and segmentation via convolutional neural network and graph cut. Int J Comput Assist Radiol Surg 2017; 12: 171–182

[43] Dror IE. Perception is far from perfection: the role of the brain and mind in constructing realities. Behav Brain Sci 2005; 28: 763–763

[44] Dror IE, Busemeyer JR, Basola B. Decision making under time pressure: an independent test of sequential sampling models. Memory & cognition 1999; 27: 713–725

[45] Verloh N, Haimerl M, Rennert J et al. Impact of liver cirrhosis on liver enhancement at Gd-EOB-DTPA enhanced MRI at 3 Tesla. European journal of radiology 2013; 82: 1710–1715

[46] Verloh N, Haimerl M, Zeman F et al. Assessing liver function by liver enhancement during the hepatobiliary phase with Gd-EOB-DTPA-enhanced MRI at 3 Tesla. Eur Radiol 2014; 24: 1013–1019

[47] Verloh N, Utpatel K, Haimerl M et al. Liver fibrosis and Gd-EOB-DTPA-enhanced MRI: A histopathologic correlation. Scientific reports 2015; 5: 15408

[48] López-Mir F, Naranjo V, Angulo J et al. Liver segmentation in MRI: a fully automatic method based on stochastic partitions. Comput Methods Programs Biomed 2014; 114: 11–28

[49] Bereciartu A, Picon A, Galdran A et al. Automatic 3D model-based method for liver segmentation in MRI based on active contours and total variation minimization. Biomed Signal Process Control 2015; 20: 71–77