Neurocomputing 505 (2022) 388–401

Contents lists available at ScienceDirect

Neurocomputing

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

Adaptable volumetric liver segmentation model for CT images using region-based features and convolutional neural network

Vanda Czipczer [b], Andrea Manno-Kovacs [a,*]

[a] Machine Perception Research Laboratory, Institute for Computer Science and Control (SZTAKI), Eötvös Loránd Research Network (ELKH), H-1111 Kende u. 13-17, Budapest, Hungary
[b] Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, H-1083 Práter utca 50/A, Budapest, Hungary

ARTICLE INFO

Article history:
Received 24 September 2021
Revised 13 May 2022
Accepted 12 July 2022
Available online 16 July 2022
Communicated by Zidong Wang

Keyword:
Liver segmentation
Convolutional neural networks
3D U-Net
Biomedical volumetric image segmentation
Fully automated segmentation

ABSTRACT

Liver plays an important role in metabolic processes, therefore fast diagnosis and potential surgical planning is essential in case of any disease. The automatic liver segmentation approach has been studied during the past years and different segmentation techniques have been proposed, but this task remains a challenge and improvements are still required to further increase segmentation accuracy. In this work, an automatic, deep learning based approach is introduced, which is adaptable and it is able to handle smaller databases, including heterogeneous data. The method starts with a preprocessing to highlight the liver area using probability density function based estimation and supervoxel segmentation. Then, a modification of the 3D U-Net is introduced, which is called 3D RP-UNet and applies the ResPath in the 3D network. Finally, with liver-heart separation and morphological steps, the segmentation results are further refined. Segmentation results on three public databases showed that the proposed method performs robustly and achieves good segmentation performance compared to other state-of-the-art approaches in the majority of the evaluation metrics.

1. Introduction

Liver plays an important role in metabolic processes, therefore fast diagnosis and potential surgical planning is essential in case of any disease. To achieve this, usually abdominal Computed Tomography (CT) scans are acquired, therefore a high amount of diagnostic data should be processed quickly. Computer aided diagnosis is often required to accelerate the process, where accurate liver segmentation is one of the main steps. Automatic liver segmentation has been studied during the past years and different techniques have been proposed, but improvements are still required to further increase segmentation accuracy, as the task is very complex. Beside the large variability of the liver's shape and size, potential pathological changes inside the organ and heterogeneous image data with different characteristics also mean a great challenge.

In the past few years, semi-automatic and fully automatic liver segmentation methods were mainly published [1]. These segmentation approaches could be divided into three main groups: image based, shape prior based and learning based techniques.

Image- (or pixel-) based approaches mainly exploited the typical image characteristics of the liver region first, using a priori knowledge, like an atlas-based registration step [2] to determine the potential organ area. This was followed by various classification steps for segmenting the final liver area, such as thresholding, region growing [3], edge detection [4], some kind of atlas-based or graph-based method [5] or active contour based outline detection [6,7].

In [8], a complex image based segmentation process was introduced for surgical planning, also extracting liver vessels and classified liver segments. Similarly to our introduced method, a histogram analysis was applied to define the potential liver area. The method was quite complex and it was quite accurate, however, required higher computational time (average liver segmentation time was 55 s). In a similar manner, [9] used region growing and level-set active contour together with a novel signed pressure function integrating both local and global information. Despite its high performance, the method required user interaction for initial seed point definition, which limited its usability in case of large databases.

Graph cuts were also widely used tools for liver segmentation, as they could handle regional and shape image features and boundary regularization [10,11]. In earlier works, human intervention was required for initialization [12,13], but later different

* Corresponding author.
E-mail addresses: vanda.czipczer@sztaki.hu (V. Czipczer), andrea.manno-kovacs@sztaki.hu (A. Manno-Kovacs).

https://doi.org/10.1016/j.neucom.2022.07.024
0925-2312/

automation steps were proposed, either estimating the potential liver location [14] or using neighboring organs (like the heart and the kidneys) [5] to select foreground/background points automatically.

The *shape-based* models, such as [15], required an average shape model, which was typically built using a larger database. The disadvantage of such models was their adaptability to large deformations. To eliminate this, a statistical shape model was introduced in [16], which combined learned shape priors with different local constraints for robust model initialization, followed by a multi-tiered model adaptation. The adaptation process was further improved in [17], where potential boundary points were classified into genuine and dubious groups, and the latter ones were then refined. A more challenging issue for these models was the liver segmentation in heterogeneous data with different modalities, in [18], a robust multi-variability 3D model for CT and Magnetic Resonance Imaging (MRI) modalities was introduced, built on the liver shape variance.

Pixel-based and shape-based elements can be fused, as in [19], where the initial liver region was detected using atlas label propagation and fusion, followed by a graph cut framework. Similarly, in [20] the authors fused a novel level set framework for initial segmentation, sparse shape composition for refinement and graph cut for further optimization. Although the method's performance was outstanding, the average running time for one CT volume was 16.6 minutes, which made real application difficult.

As our aim was to propose a fast method even for smaller databases including heterogeneous data (i. e. scans from different hospitals in a real life application), shape models with large shape variability could not be built, therefore we preferred image-based methods throughout our work.

Learning-based methods were lately introduced for medical image analysis and attracted great attention, as they had the ability to extract multiple features and to make segmentation more accurate. Different neural networks were introduced for medical image analysis, including Convolutional Neural Networks (CNNs) [21,22], Fully Convolutional Neural network (FCN) [23–26], Deep Belief Network [27], etc. Probably, one of the most popular CNN architecture is the U-Net model [21], which was published for medical image analysis in different application fields. Since then, many modifications of the original network were introduced for various applications [28,29,22,23,30]. Recently, the U-Net architecture was also used for liver segmentation [31–34], however, the majority of the proposed modifications performed a slice-by-slice 2D segmentation, which made the computational time higher and did not exploit potential interslice information. Learning-based models were also fused with pixel-based techniques for refining the segmentation result, like in [23], where a 3D deeply supervised network was proposed, and the output probability mask was further refined with a fully connected conditional random field model (CRF). Similarly, the CRF was also applied in [24] for refinement, after using cascaded fully convolutional neural networks for liver and lesion segmentation. Active contour method was used for refinement in [27], where deep belief network was combined with deep neural network to perform a complex training in two steps. In a recent work [35], improved Faster R-CNN [36] was applied for approximating the liver location as a first step, then a liver segmentation with DeepLab in the second stage.

The latest advances in medical image processing applications are the transformer-based architectures. Self-attention-based architectures, in particular Transformers [37], are the leading model in natural language processing [38]. Their dominant approach is two-step: first, a pretraining on a large text corpus, then a fine-tune on a smaller task-specific dataset. Inspired by their success, there is a continuous effort to introduce them also in computer vision. Either combining them with CNN-like architectures [39], or replacing convolutions entirely [40]. Transformer-based architectures are also introduced in medical image processing, following mainly the fusion model by combining a Transformer into the U-Net architecture [41–44]. Usually, these models are trained on larger data sets and require huge hardware. They prefer training databases where either the major part of the image is annotated (such as the BTCV Synapse Multi-organ CT Database or the MSD MRI/CT database), or the database itself is quite huge (like the MICCAI BRATS challenge) to have enough training data to learn the context. Published liver segmentation methods are also evaluated on the mentioned multi-atlas databases (the BTCV and MSD), and the results are quite promising (DSC values from 92.2 to 98.5 for the BTCV, and 95.4 for the MSD). However, no segmentation results are available at this point for smaller datasets, like the ones we used in our evaluation, therefore, it is impossible to give a fair quantitative comparison with those models.

In our work, we were inspired by the 3D extension of the original 2D U-Net (called 3D U-Net) [22] and the modification proposed by [30], where the MultiRes block and the ResPath were introduced. Unlike [45], which adapted the 2D ResUNet followed by a 3D postprocessing step, we have integrated the ResPath into a 3D U-net based network for a one-step volumetric segmentation, which is called the 3D RP-UNet.

As a novel contribution, we have introduced an end-to-end 3D segmentation approach for heterogeneous CT data with improved pre- and postprocessing steps, resulting in an accelerated, yet efficient processing strategy even for smaller databases with different image characteristics. The proposed preprocessing runs completely in 3D and emphasizes the liver region by using probability density function and supervoxel segmentation. After the learning-based volumetric segmentation, some typical segmentation problems are corrected by the proposed postprocessing, including a liver-heart separation step. Beside the high segmentation accuracy and the low computation cost, the introduced approach is also capable of generalization and handles previously unseen data efficiently; therefore it has the potential to be used in real life applications as well.

Evaluation results on three publicly available datasets - MICCAI Sliver07 [1], the 3D-IRCADb [46] and the VISCERAL Anatomy3 database [47] - show that the proposed method has strong performance compared to the state-of-the-art approaches, regarding the majority of the calculated six evaluation metrics.

2. Methodology

2.1. Preprocessing

We have improved our method, which was previously proposed in [48] and motivated by [5] to run completely in 3D. The preprocessing of the CT scan starts with a rescaling, based on its maximum Hounsfield unit value and the rescaled scan is then stored in a 16-bit image (Fig. 1(b)). This way we acquired a wider range of pixel intensities. After rescaling, to reduce computational complexity, the volumes are resampled to $128 \times 128 \times 64$. For further processing of the CT scan, median filtering is used to reduce noise while preserving edges, which is important for image segmentation [4].

Following [48], our presumption is that the liver is located in the upper half of the volume (considering the orientation of the slices shown in Fig. 1). To verify the validity of this presumption for the actual CT scan, we checked the location of the spine using the axial maximum intensity projection (MIP) of the bone mask, inspired by [5]. If the spine is detected on the left side of the image (thus, the liver is located in the lower half of the volume), it means

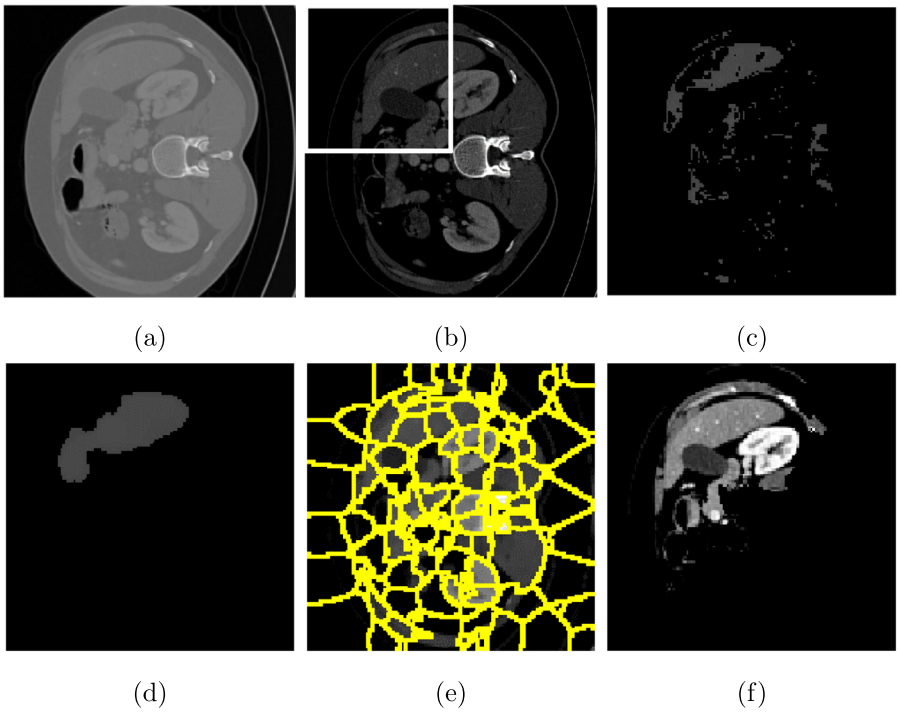


Fig. 1. The steps of preprocessing: (a) Original image, (b) Rescaled image (white rectangle indicates the image quarter where the liver is assumed), (c) Thresholded image, (d) Morphologically modified image, (e) Supervoxel regions of the rescaled image, (f) Preprocessed image.

that the whole scan is rotated. Therefore, the volume have to be rotated with 180 degrees before further steps.

The image is then thresholded using the probability density function (PDF). The input of the PDF estimation is one of the upper quadrants of the volume (marked by a white rectangle in Fig. 1(b) for one slice), depending on whether the scan is slightly rotated or not. The degree of rotation was determined in the previous step, using the axial MIP of the bone mask. Therefore, we expect that the selected quadrant contains most of the liver voxels. To further highlight the liver region, we also excluded the voxels with intensity lower than 500 (based on our previous experiments), thus, one of the PDF peaks is expected to represent the intensity range of the liver.

Based on the analysis of the calculated PDFs, the CT scans could be classified into two categories: high contrast or low contrast (Fig. 2) [5,48]. The high contrast scans have more than one peak in their PDF (right image of Fig. 2), and the second peak represents the liver intensity range. The low contrast scans have only one peak (left image of Fig. 2), which corresponds to the intensity range of the liver. Therefore, the boundaries of the thresholded intensity range are defined by the location and the width (half-height) of the corresponding peak. This thresholding method removes most of the voxels of the other organs (as shown in Fig. 1(c)). The output of this step is further refined by previously proposed cavity filling and morphological opening. As an improvement, we further extended these refinement steps with keeping the largest connected component and a morphological dilation. The output of these steps is shown in Fig. 1(d).

The obtained, thresholded volume may not include some liver parts such as large vessels and lesions next to the boundary due to higher or lower intensity, respectively, relative to the liver intensity range. We propose 3D supervoxel segmentation to retrieve

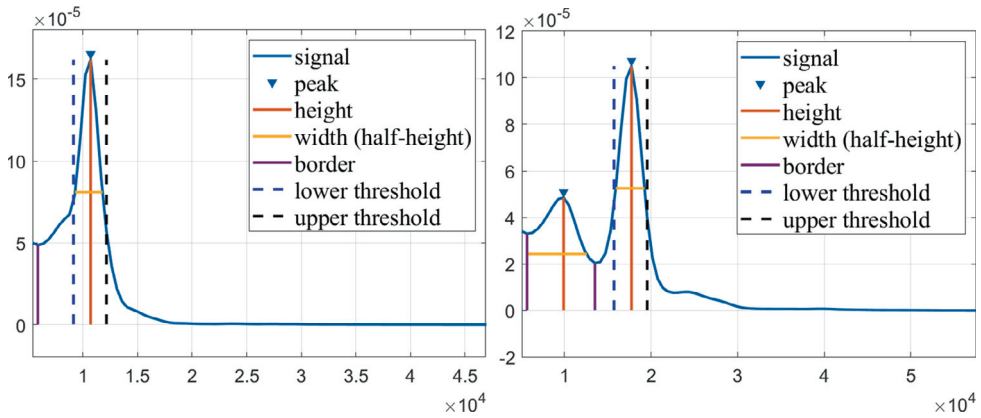


Fig. 2. Probability density functions (PDFs) of the pixel intensity values. Left image is an example for the PDF of a low contrast scan; right image is an example for the PDF of a high contrast scan. Blue dash line represents the lower boundary, while the black represents the upper boundary of the estimated liver intensity range.

these falsely removed liver parts. The supervoxel algorithm is based on the modified 3D version of the Simple Linear Iterative Clustering (SLIC) algorithm [49]. SLIC is a novel k-means based clustering algorithm, which can generate superpixels/ supervoxels quickly and efficiently based on the intensity and the 2D/3D spatial proximity. A weighted distance measure combines these features, while simultaneously providing control over the size and compactness of the superpixels or supervoxels. As a result, the calculated supervoxels have nearly uniform size, while their boundaries closely match the true image edges.

In our work, we used the built-in Matlab function *superpixels3* for computing 3D supervoxels. The function requires the 3D volume and the N_d desired number of supervoxels as input. It returns the N_a actual number of the calculated supervoxels together with a 3D label matrix. Fig. 3 shows the output of the clustering for different N_d numbers of supervoxel regions. As it can be seen, when the size of the supervoxel region is smaller (thus the N_d number of the supervoxels is larger), the supervoxel algorithm can find the liver border more precisely (Fig. 3(c)).

The supervoxel segmentation is applied to the rescaled and resampled CT scan (Fig. 1)) and the refinement is performed by removing only those supervoxel regions, that don't contain any voxel of the obtained thresholded volume. We have to set the N_d number of supervoxel regions to relatively small, so that the supervoxel regions include those possibly removed liver parts, but if we set the N_d number to too small, the regions include unnecessary non-liver tissues. Based on our experiments, the N_d number is set to 410, but the N_a actual number of the created supervoxel regions is between $[250, 350]$.

Lastly, the remaining image parts are rescaled again to enhance the image contrast. As Fig. 1(f) shows, most of the non-liver tissues are removed and the remaining organs can be more easily separated than in the original image (Fig. 1(a)). This preprocessed image will be the input of the proposed CNN model.

2.2. 3D RP-UNet

The neural network, which was implemented in this work, is a modified version of the 3D U-Net [22] (Fig. 4), which we call the 3D RP-UNet. In the 3D RP-UNet (Fig. 5), we use upsampling in the expanding path, instead of up-convolution (marked by green upwards arrows) to reduce the number of total parameters and to have only trainable parameters. Furthermore, we exchanged the skip connections with the ResPath, introduced in [30]. The skip connection is a simple concatenation of the feature maps of the corresponding layers of the contracting and expanding stage. This enables the network to propagate spatial information lost during the pooling operations. Ibtehaz and Rahman states [30] that there is a flaw of the skip connections, since there is a semantic gap between the corresponding layers of the encoder-decoder stage. Therefore, the ResPath was proposed to resolve this problem and instead of a simple concatenation, the feature maps from encoder stage are passed through the ResPath, and then concatenated with the decoder features. The ResPath is a chain of convolutional layers with residual connections (illustrated in Fig. 6). The residual connections - earlier used for example in [50] - are introduced into the ResPath since they increase the convergence speed, reduce the number of convolutional blocks and they are also proven to have great potential in the medical image analysis [51].

As a novelty, we introduced the ResPath into a 3D U-Net network (Fig. 5), so we used $3 \times 3 \times 3$ convolutional layers and $1 \times 1 \times 1$ residual connections. Additionally, the 3D RP-UNet consists of 3 ResPaths with $3, 2, 1$ convolutional blocks, respectively

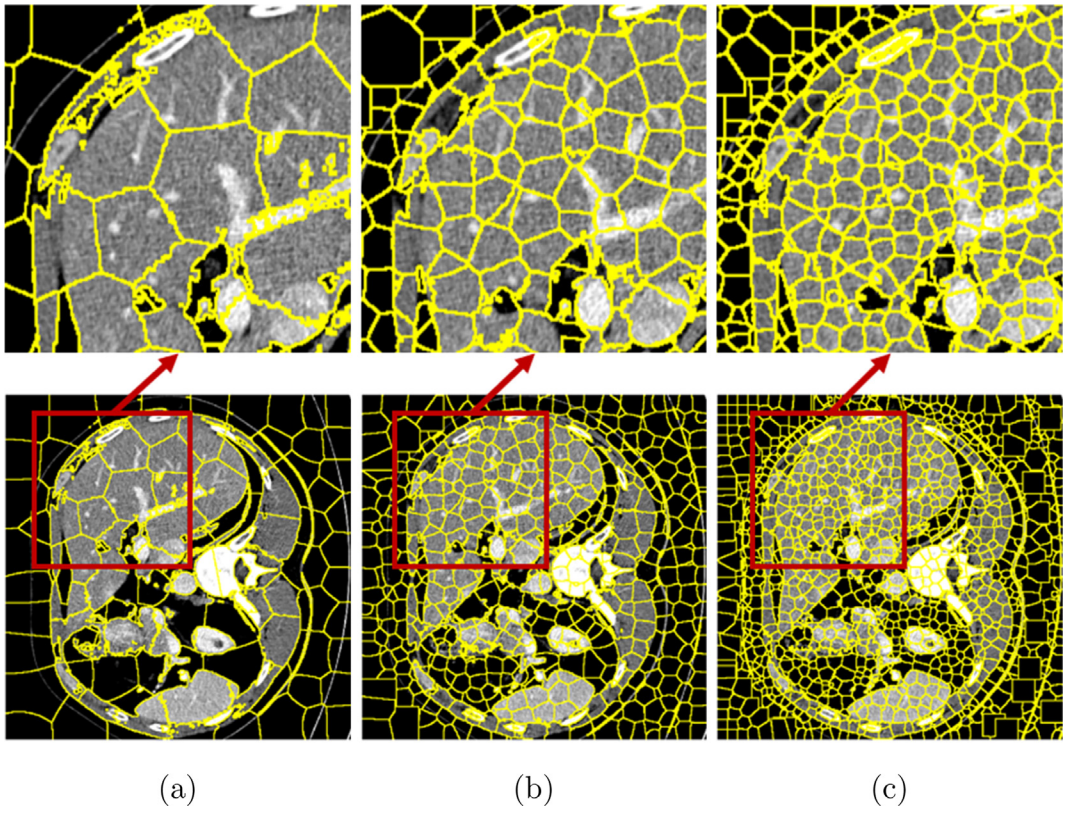


Fig. 3. Examples of supervoxel segmentations produced by different N_d numbers of supervoxels on one CT slice. The different numbers of supervoxels, from (a)-(c), is $100, 1000, 5000$, respectively. The image data inside the red box in (a)–(c) are displayed in the top row.

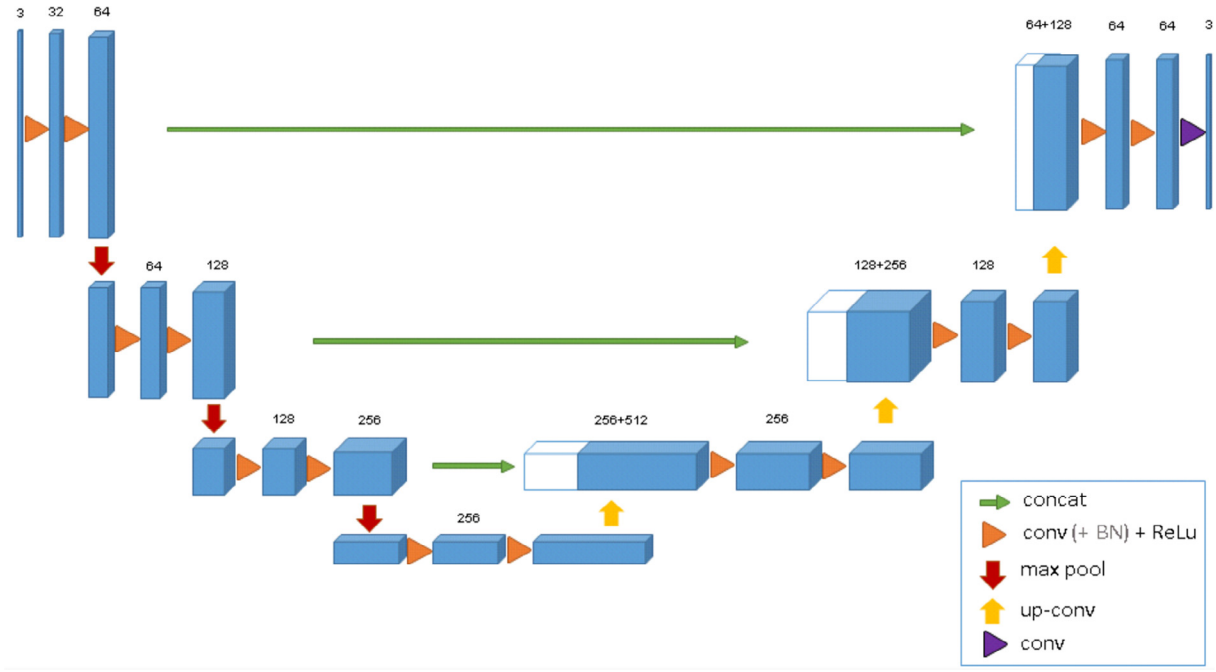


Fig. 4. The architecture of 3D U-Net in [22]. Blue boxes represent the feature maps obtained from the convolution operations. The number above them represents the number of feature maps.

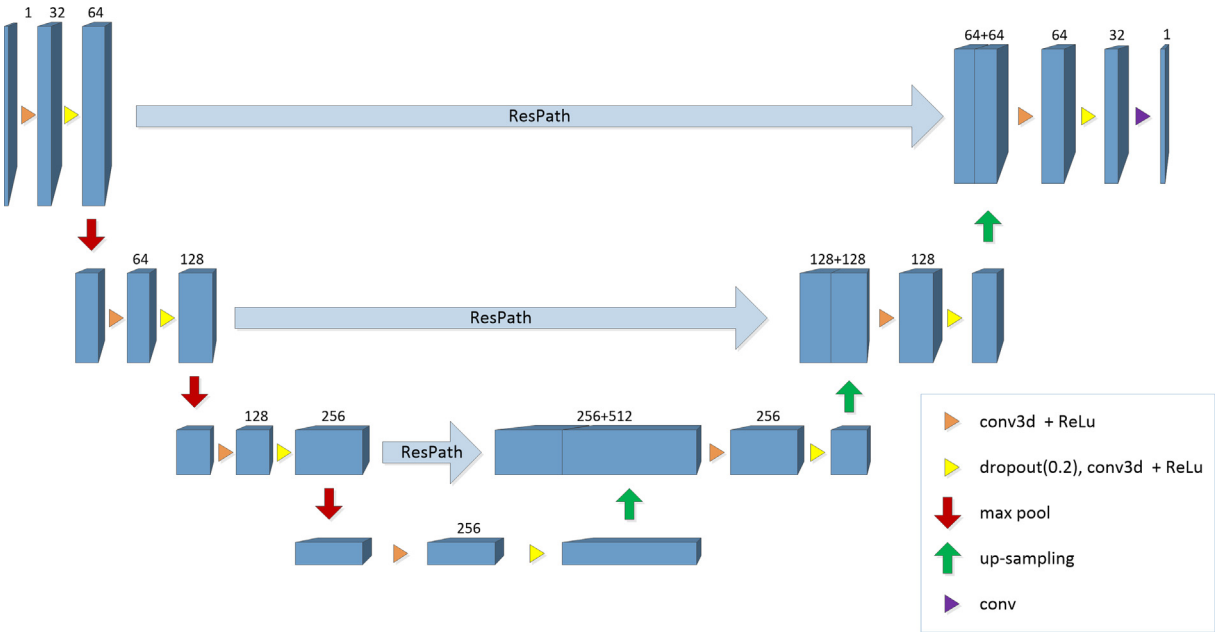


Fig. 5. The architecture of the proposed 3D RP-UNet is an improvement of 3D U-Net [22]. Blue boxes represent the feature maps obtained from the convolution operations. The number above them represents the number of feature maps.

and the number of feature maps used in the blocks of the ResPath are $64, 128, 256$ respectively. Furthermore, we used dropout layer with $p = 0.2$ between the $conv3d$ functions along the U-shape, and skipped the Batch Normalization of the original 3D U-Net to reduce GPU memory. Moreover, we changed the number of feature maps in the expansive path due to limited GPU memory, thus the number of feature maps in the proposed model is $256, 128$ and 64 before the up-sampling functions.

2.3. Postprocessing

The output of the neural network is a probability map with values between 0 and 1. To get a binary prediction mask of it, the probability map was thresholded at the value of 0.5. In the obtained binary prediction masks, in some cases, small oversegmentation could be observed at the heart (Fig. 7(a)) or the spleen (Fig. 7(b)) (if the preprocessing wasn't so accurate) and underseg-

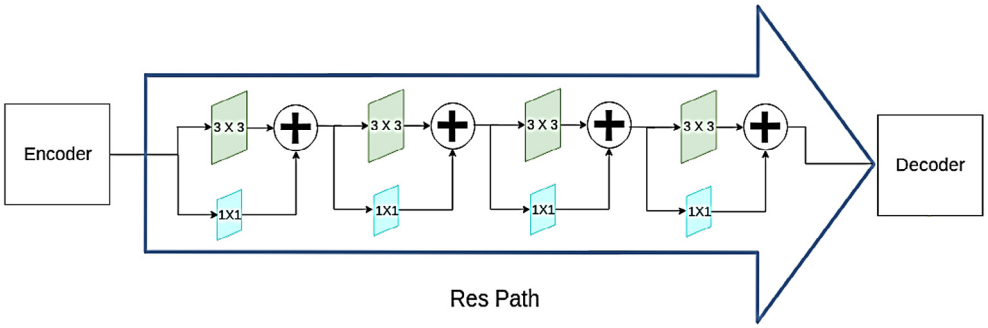


Fig. 6. ResPath proposed in [30].

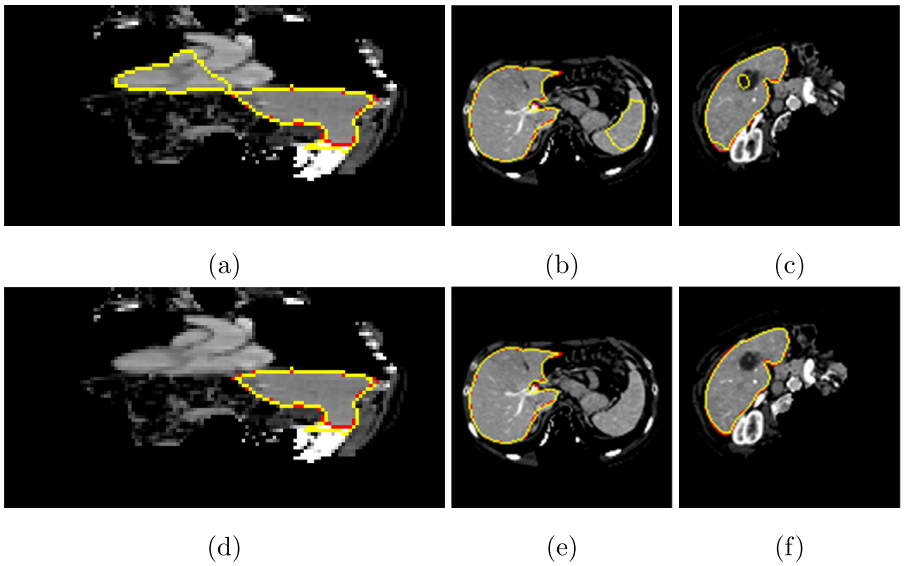


Fig. 7. Improvement of the prediction masks presented on preprocessed images: (a) and (b) Oversegmentation in the place of (a) heart and (b) spleen and (c) undersegmentation inside the liver due to lesion. (d) - (f) Elimination of the missegmentation. Red contour represents the ground truth, while yellow contour shows the segmentation results.

mentation because of a lesion inside the liver (Fig. 7(c)). Thus, the postprocessing focused on these problems.

First, we applied the liver-heart separation method, introduced in [3] which was improved and accelerated in our previous work [48]. Our improved version first determines the binary mask of the air-filled parts of the 3D abdominal region with a thresholding step (a sample 2D slice is shown in Fig. 8(b)). The orientation of the volume is assumed to be vertical and the lungs are connected to the upper border (as shown in Fig. 8). Then, all the components of the obtained 3D binary volume are categorized into the following classes:

1. Components connected to left, right or bottom border of the 3D abdominal region (marked in red in Fig. 8(c));

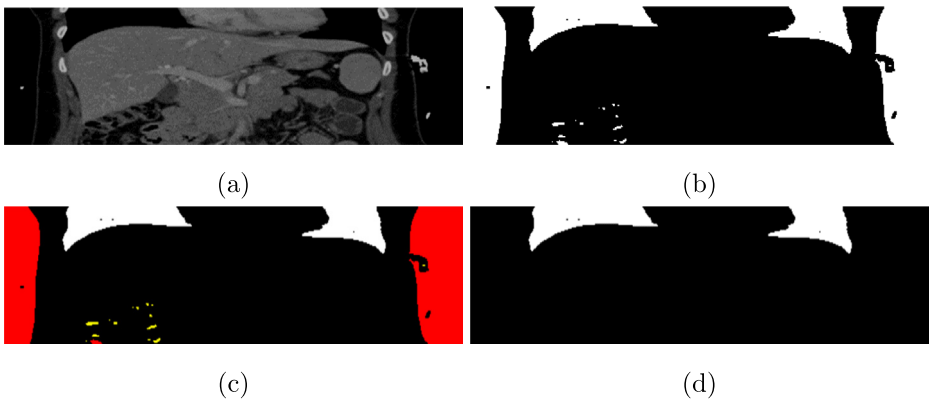


Fig. 8. Illustration of the lung segmentation on a sample coronal slice: (a) Original image; (b) Obtained binary mask of air-filled parts; (c) Filtered regions (yellow and red blobs); (d) Result of the segmentation.

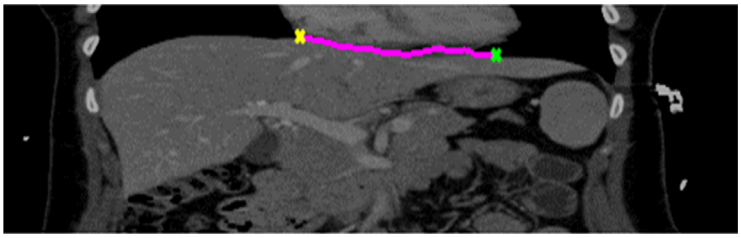


Fig. 9. The separating curve between the heart and liver. Leftmost point is presented in green and the rightmost point in yellow.

2. Components not connected to any borders (yellow in Fig. 8(c)).
3. Components not categorized into the previous two classes (white in Fig. 8(c)).

After the classification, the components categorized into the first two classes (marked in red or yellow in Fig. 8(c)) are eliminated. The remaining components in the 3D binary mask are expected to represent the lung mask (Fig. 8(d)), which is applied for separating the liver and the heart. A curve is defined between the right and left lobe for each coronal slice. The curve starts from the rightmost point of the patient's left lobe (marked in green in Fig. 9) and from right to left, each point of the curve is chosen based on the location of the largest gradient value found in the local environment of the previous point. The process is finished when it reaches the leftmost point of patient's right lobe (marked in yellow in Fig. 9). The set of these curves defines a surface, which connects the two lobes and separates the heart (above the surface) and the liver (under the surface) [3]. Lastly, the curves located in neighboring slices are averaged to provide a smoother separating surface.

Above the surface defined by the liver-heart separation, we set every voxel of the binary prediction mask to zero. Then, we only keep the largest connected component to remove the non-connecting blobs such as the one in Fig. 7(b). In the next step, morphological closing operation is applied to smooth the surface of the predicted liver mask and we only keep those voxels that is inside the original thresholded probability map. Lastly, cavity filling is applied on the binary mask on each axial slice to fill in the holes caused by the potential lesions inside the liver (Fig. 7(c)). This step gives the final liver mask.

3. Results

3.1. Datasets

During this work, we used three different, publicly available databases: the MICCAI Sliver07 [1], the 3D-IRCADb [46] and the VISCERAL Anatomy3 database [47].

The MICCAI Sliver07 database contains 20 clinical contrast-enhanced (portal venous phase) abdominal CT volumes acquired using a variety of scanners, including ground truth images with expert segmentation. Most of the data is pathological. The number of slices, in-plane resolution, and interslice resolution varies between 64 and 394, 0.58 and 0.81 *mm*, and 0.7 and 5.0 *mm*, respectively. 3D-IRCADb database is also composed of 20 venous phase enhanced CT scans with hepatic tumors in 75% of the cases. On the IRCAD website, a detailed description is added for each scan, including liver size (width, depth, height) or the liver's average density. It also lists what could cause segmentation errors, including the contact with neighboring organs, an atypical shape or the density of the liver or even artifacts in the image. The number of slices, in-plane resolution, and interslice resolution ranges between 74 and 260, 0.561 and 0.873 *mm*, and 1.0 and 4.0 *mm*, respectively.

The VISCERAL Anatomy3 database includes 79 samples with manual liver segmentation, from which we only use the 20 contrast enhanced CT images. It differs from the above-mentioned 2 databases, as all the CT images include the thorax and each liver in the database is non-pathological.

The segmented liver regions were defined to include the entire liver tissue and all internal structures, including vessels, tumors, and cirrhosis.

3.2. Evaluation measures

The results of the different methods were quantitatively evaluated using 6 different metrics by comparing them to the manual segmentation done by experts. The 6 metrics, often used in the state-of-the-art, are the followings:

Dice similarity coefficient (DSC) measures the similarity of two binary image:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{1}$$

Volumetric overlap error (VOE) [%]:

$$VOE = 100\left(1 - \frac{A \cap B}{A \cup B}\right) \tag{2}$$

Relative volume difference (RVD) [%]:

$$RVD = 100\left(\frac{|A| - |B|}{|B|}\right) \tag{3}$$

Average symmetric surface distance (ASD) [mm]:

$$ASD = \frac{1}{|S(A)| + |S(B)|}\left(\sum_{S_A \in S(A)} d(S_A, S(B)) + \sum_{S_B \in S(B)} d(S_B, S(A))\right) \tag{4}$$

Root mean square symmetric surface distance (RMSSD) [mm]:

$$RMSSD = \sqrt{\frac{1}{|S(A)| + |S(B)|}\left(\sum_{S_A \in S(A)} d^2(S_A, S(B)) + \sum_{S_B \in S(B)} d^2(S_B, S(A))\right)} \tag{5}$$

Maximum symmetric surface distance (MaxD) [mm]:

$$MaxD = \max\left\{\max_{S_A \in S(A)} d(S_A, S(B)), \max_{S_B \in S(B)} d(S_B, S(A))\right\} \tag{6}$$

In DSC metric, true positive (TP) means, the observed value is 1 and is predicted to be 1. False positive (FP) means that the observed value is 0, but it is predicted to be 1. False negative (FN) is the opposite of FP: the observed value is 1, but it is predicted to be 0.

In the last five measures, A represents the segmented volume produced by the proposed method, B is the ground truth, $S(A)$ denotes the set of surface voxels of A. The shortest distance of a voxel v to $S(A)$ is defined as $d(v, S(A)) = \min_{S_A \in S(A)} \|v - S_A\|$, where $\|.\|$ represents the Euclidean distance [5]. DSC is 1, the other metrics are zero for a perfect segmentation.

3.3. Data augmentation

For data augmentation, we adopted 3 techniques for all training data to increase the training dataset, thus alleviating the overfitting problem:

- rotation of images by 5 to 75 degree,
- random shearing by 5 to 20 degree,
- and scaling images to a value of 80% to 120% of the original size.

Each augmentation step is performed only in the axial direction.

3.4. Ablation study

We compared the performance of the following different models:

1. the original 3D U-Net [22] with 19078337 parameters in total (including 4672 non-trainable parameters);
2. 3D U-Net without upconvolution (marked by "w/o upconv") to reduce the number of total parameters (16315585 parameters in total, only trainable parameters);
3. the second model with added batch normalization (marked by BN) to avoid overfitting;
4. the proposed 3D RP-UNet (17699137 parameters in total, only trainable parameters).

We trained these implemented models 10 times on the same set of MICCAI Sliver07 database with the same hyperparameter settings. The training dataset contained 18 volumes of the MICCAI Sliver07 database. We applied only a minimal preprocessing to the volumes. First, to reduce computational complexity, all volumes were resampled to $[128, 128, 64]$, then the intensity range of all the volumes is normalized to $[-128, 128]$. Therefore, the performance of each model is evaluated without a more complex pre- and postprocessing.

Throughout the training, each batch contained only one resized 3D volume of CT images. The number of epochs was set to 165 with 18 steps per epoch. The chosen optimizer was Adam with the initial learning rate of 10^{-4}. The performance of the models was measured on each epoch using the DSC and binary cross-entropy loss.

Fig. 10 shows the average performance using the DSC metric, Fig. 11 shows the average loss of the models. It can be noted, that the proposed 3D RP-UNet attains convergence a bit faster than the other 3D U-Net versions, therefore we are likely to obtain better results in a smaller number of training epochs as compared to the other models. Similarly, 3D RP-UNet achieves lower validation loss, than the other 3D U-Net based models. By comparing the 3D U-Net based models, 3D U-Net w/o upconv with BN converges a bit slower, than the other two compared versions.

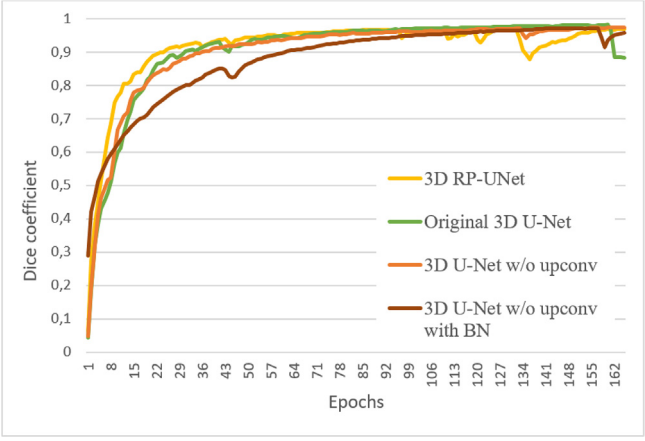


Fig. 10. Progress of the average training performance for the different U-Net versions. Upconv is the abbreviation of the upconvolution, and BN marks the batch normalization.

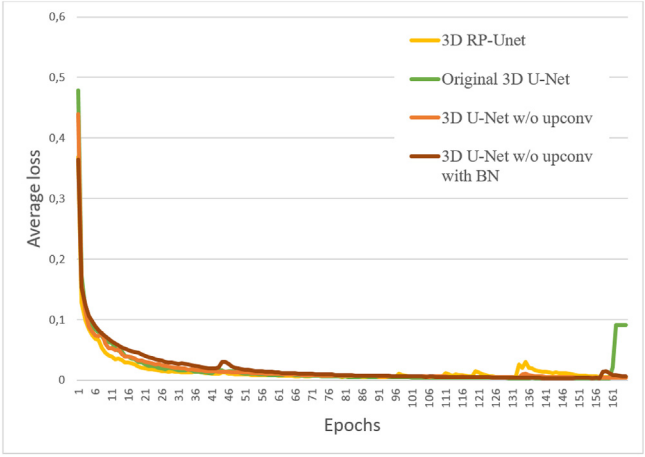


Fig. 11. Progress of the average training loss for the different U-Net versions. Upconv is the abbreviation of the upconvolution, and BN marks the batch normalization.

To further analyze the performance quantitatively, the models' results were also evaluated on two test volumes using the 6 metrics, introduced in Section 3.2. Throughout the evaluations, we used the models with weights that gave the best DSC during training. Since each model's result was a probability map with values between 0 and 1, we used a 0.5 threshold to get a binary mask of the liver and we compared this mask to the ground truth.

In Table 1, the best average results obtained from the quantitative evaluation are shown in bold. It can be seen that the 3D U-Net w/o upconv with BN and the 3D RP-UNet overperformed the other two models and achieved similar evaluation results. Due to the added batch normalization, 3D U-Net w/o upconv with BN could generalize more easily than the original 3D U-Net and 3D U-Net without upconvolution. When comparing the two better performing models, their performance is quite similar for all metrics in case of the test1 volume, however for the test2 scan 3D RP-UNet performs significantly better considering the RVD, RMSSD and MaxD metrics.

Considering these observations in addition to the fact that 3D RP-UNet is the best converging model, we can state that 3D RP-UNet had the best performance in the ablation experiments and together with the pre- and postprocessing steps the performance could be further improved.

3.5. Training, Validation and Testing

The implemented 3D RP-UNet is trained and validated on the preprocessed and augmented CT images, and tested on the merely preprocessed CT images. Throughout the training, each batch contained only one 3D CT scan. The number of epochs was set to 100 and the chosen optimizer was Adam with the initial learning rate of 10^{-4}. We measured the performance of the models on each epoch using the DSC and binary cross-entropy loss.

Additionally, we stored the weights of the model that gave the best validation DSC during training, and throughout the evaluations, we used those prediction masks, that were obtained by the model with these weights.

Table 1
Quantitative performance of the different 3D U-Net models on 2 test images from the MICCAI Sliver07 database.

		DSC	VOE	RVD	ASD	RMSSD	MaxD
Original 3D U-Net [22]	*test1*	0.96	8.23	−**0.22**	0.70	2.32	26.61
	test2	0.93	13.55	3.77	0.56	**1.37**	**10.23**
3D U-Net	*test1*	0.95	9.00	0.47	0.85	2.84	29.38
w/o upconvolution	*test2*	0.93	12.75	**0.38**	0.83	2.62	19.27
3D U-Net	*test1*	**0.97**	**6.45**	−0.98	**0.61**	2.38	28.59
w/o upconvolution with BN	*test2*	**0.94**	**10.49**	5.07	0.63	2.34	24.66
3D RP-UNet	*test1*	0.96	6.94	−0.90	**0.61**	**2.14**	**21.81**
	test2	**0.94**	11.82	−0.55	**0.53**	1.61	14.96

3.6. Experiments and evaluation

The proposed method was trained and tested on the MICCAI Sliver07 and 3D-IRCADb databases, and only tested on the VISCERAL Anatomy3 database (to evaluate the performance also on previously unseen data). The first step was the preprocessing of the CT volumes to remove most of the non-liver tissues. The average computation time for a whole CT scan took 8.4 seconds, measured on randomly selected volumes of the MICCAI Sliver07 database. In most cases, the preprocessing worked efficiently. However, in one extreme case, small liver area was removed due to a large lesion with distinctly different intensity values next to boundary (Fig. 12(a)). Additionally, in a few cases (5 out of 60), relatively small liver tissues were also removed at the liver boundary (example shown in Fig. 12(b)).

The preprocessed and then augmented CT images were then the input for the 3D RP-UNet.

The 3D RP-UNet was trained and tested on each database using the leave-one-out method, which means we only test on one volume; the rest will be the training set. Before the training, the training set was augmented as described in Section 3.3, then randomly divided into training dataset (90%) and validation dataset (10%). During training, the hyperparameter settings remained the same as it was described in Section 3.5. The average prediction time for one CT volume of the MICCAI Sliver07 database took 6.2 seconds. Lastly, the obtained prediction masks were postprocessed. The average computational time of the postprocessing was 1.2 seconds for randomly selected CT scans from the MICCAI Sliver07 database.

The proposed method's results are quantitatively evaluated using the 6 metrics (Section 3.2) and are compared to the state-of-the-art. The quantitative results are shown in Table 2 for the MICCAI Sliver07 database and in Table 3 for the 3D-IRCADb database.

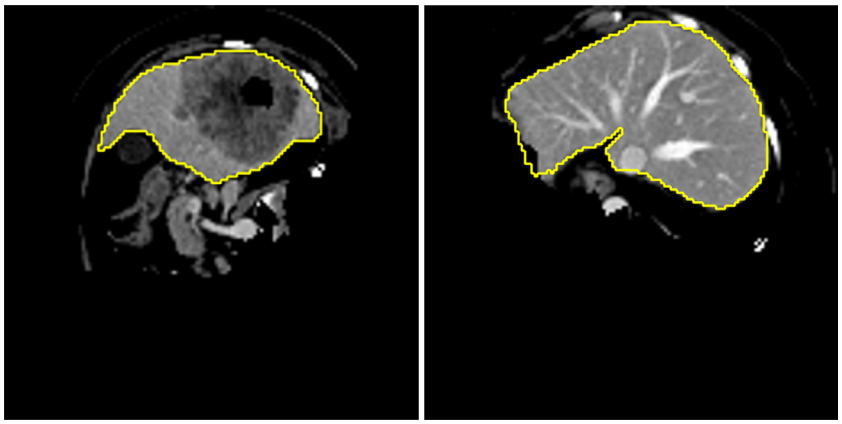
The evaluation metrics show that for the MICCAI Sliver07 database, our method achieves average VOE, however for the other metrics the performance is quite strong. Further discussion of the results is given in Section 4.

In case of the 3D-IRCADb dataset, the achieved VOE is also average, the DSC is among the higher values. The RVD is the second best along with the state-of-the-art method proposed in [20]. Concerning all the other evaluation values (ASD, RMSSD, MaxD) of our proposed method, they are significantly better than the compared approaches' performance.

Fig. 13 shows some representative slices of the proposed method's segmentation results for one MICCAI Sliver07 scan and one 3D-IRCADb scan. The result and the ground truth are shown in the axial, coronal, and sagittal directions respectively (ground truth in red, segmentation result in yellow). The last column of Fig. 13 includes the 3D model of each segmentation.

Based on the images in Fig. 13, it can be seen that the proposed method's results approximate the reference segmentation. The liver is successfully separated from the neighboring organs: from the heart in most cases (Fig. 13)) and from the right kidney (Fig. 13. However typical segmentation errors still exist: the proposed method was less sensitive against the tumors next to liver boundary (Fig. 14(a)). Furthermore, oversegmentation happened when the liver-heart separation wasn't working efficiently (Fig. 14)) and also when some vessel parts (that don't belong to the liver based on the reference) were included due to the cavity filling in the axial direction during postprocessing (Fig. 14(c)).

To demonstrate the efficiency of our proposed preprocessing step, another leave-one-out training was performed for the MICCAI Sliver07 database, with a more simple preprocessing, introduced in [52]. This preprocessing includes resampling the image to $128 \times 128 \times 64$, normalizing values between $[-128, 128]$ and using anisotropic diffusion to reduce noise. Table 4 shows the quantita-



(a) (b)

Fig. 12. Errors occurring in preprocessing step. Yellow contour represents the liver boundary based on reference segmentation.

Table 2
The quantitative comparative results for the MICCAI Sliver07 train dataset as mean. (Note: * denotes that the results were evaluated only on a few data)

	DSC	VOE	RVD	ASD	RMSSD	MaxD
Erdt et al. (2010) [16]	-	7.54	1.28	1.30	2.67	26.52
Dou et al. (2016) [23]	-	5.37	1.32	0.67	1.48	29.63
Esfandiarkhani and Foruzan (2017) [17]	-	8.13	0.42	1.31	2.38	21.35
Lu et al. (2018) [19]	-	5.92	1.03	1.06	1.68	12.33
Huang et al. (2018) [10]	-	5.3	−0.6	1.0	2.1	21.2
Ahmad et al. (2019) [27] *	0.948	**4.31**	1.28	-	-	-
Lebre et al. (2019) [18]	0.93	-	-	-	-	-
Tang et al. (2020) [35]	-	5.06	1.11	0.78	1.70	23.42
Li et al. (2020) [20]	-	5.1	**0.1**	0.9	1.8	19.4
Wang et al. (2021) [34]	**0.973**	5.37	−1.08	1.85	-	27.45
Proposed method	0.964	6.92	-0.36	**0.33**	**0.92**	**9.27**

Table 3
The quantitative comparative results for the 3D-IRCADb dataset as mean. (Note: * denotes that the results were evaluated only on a few data)

	DSC	VOE	RVD	ASD	RMSSD	MaxD
Erdt et al. (2010) [16]	-	10.34	1.55	1.74	3.51	26.83
Christ et al. (2016) [24]	0.943	10.7	−1.4	1.5	-	24.0
Esfandiarkhani and Foruzan (2017) [17]	-	10.39	1.48	1.66	3.68	35.8
Lu et al. (2018) [19]	-	9.21	1.27	1.75	3.95	36.17
Huang et al. (2018) [10]	-	8.6	−0.7	1.6	3.1	26.9
Li et al. (2018) [31]	0.947	10.02	**−0.01**	4.06	9.63	-
Ahmad et al. (2019) [27] *	0.918	**6.09**	5.59	-	-	-
Lebre et al. (2019) [18]	0.87	-	-	-	-	-
Tang et al. (2020) [35]	-	8.67	0.57	1.37	3.56	27.01
Li et al. (2020) [20]	-	9.2	0.5	1.6	3.1	27.2
Budak et al. (2020) [33]	**0.952**	9.05	7.03	1.43	-	19.37
Proposed method	0.945	10.16	-0.50	**0.53**	**1.29**	**10.00**

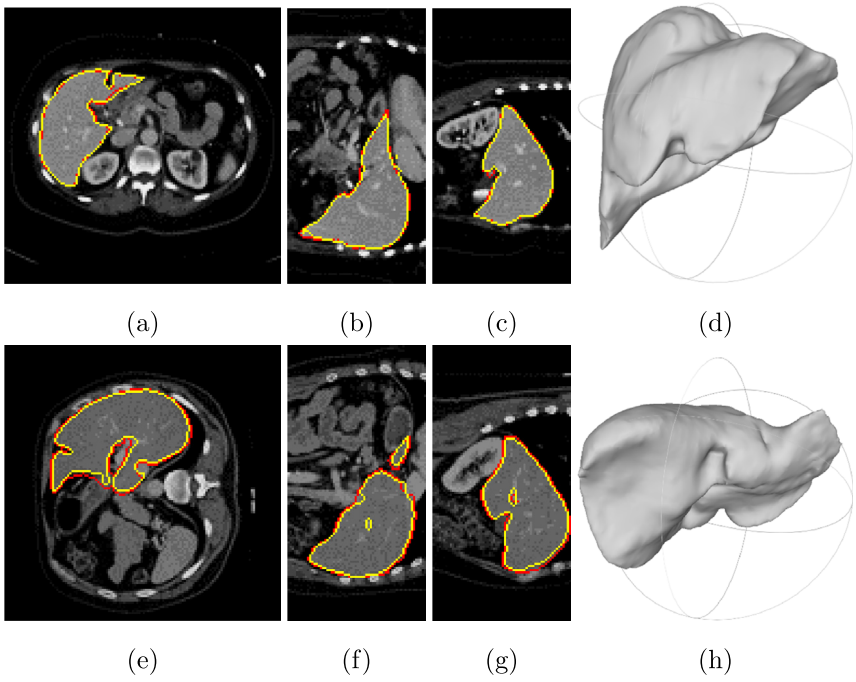


Fig. 13. Examples of liver segmentation results of the proposed method for one MICCAI Sliver07 dataset (top row) and one 3D-IRCADb dataset (bottom row). Each row shows slices of one case in the axial, coronal, sagittal directions and its 3D representation, respectively. The contour of the ground truth is in red. Yellow contour denotes the result of proposed method.

tive results of the 3D RP-UNet with different preprocessing steps without postprocessing. It should be noted, that in this case, we only used the preprocessed CT scans of MICCAI Sliver07 database for training, without any data augmentation. The better results are in bold. As it can be seen, with the proposed preprocessing, the achieved performance is better.

To show that our method is capable of generalization, we tested our model - trained on the MICCAI Sliver07 and 3D-IRCADb data-

Fig. 14. Typical errors of the segmentation method on both 3D-IRCADb and MICCAI Sliver07 database. The contour of the ground truth is in red. Yellow contour denotes the result of proposed method.

Table 4
Quantitative evaluation of the 3D RP-UNet with different preprocessing steps without postprocessing. Note that the training was performed merely on the MICCAI Sliver07 dataset without any data augmentation.

Preprocessing	DSC	VOE	RVD	ASD	RMSSD	MaxD
Method in [52]	0.936	11.809	**−0.598**	0.778	1.905	13.992
Proposed	**0.950**	**9.434**	0.705	**0.627**	**1.749**	**12.697**

Table 5
Test results obtained on the VISCERAL database with different state-of-the-art models. The first two methods are trained on different data sets and only tested on the VISCERAL; while the rest are directly trained and tested on the VISCERAL.

Method	Training database	DSC
He et al. (2021) [53]	LiTS	0.907
Proposed method	MICCAI Sliver07 + 3D-IRCADb	**0.922**
Blendowski et al. (2019) [59]	VISCERAL	0.853
Gass et al. (2014) [57]		0.908
Kéchichian et al. (2014) [55]		0.933
He et al. (2015) [56]		0.933
Oliveira et al. (2018) [58]		0.939

bases -, on the VISCERAL database. The results are listed in Table 5. In a similar manner, a study on the generalization ability of 3D U-Net was published in [53] using multiple heterogeneous liver CT datasets including VISCERAL database. The highest achieved DSC score of the generalized 3D U-Net model was 0.907 and it was trained on the LiTS database [54], including 130 abdominal CT volumes scanned in multiple clinical sites.

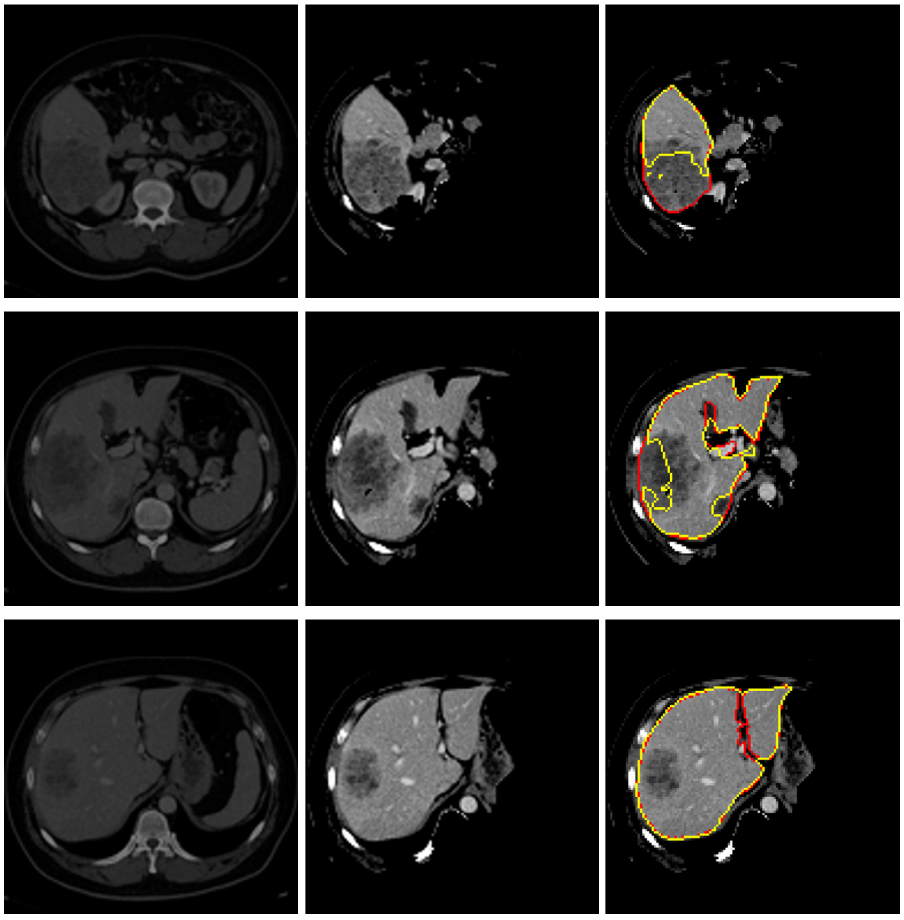
Beside, some state-of-the-art evaluation results are available directly trained and tested on the VISCERAL database itself. Such results are also added in the lower part of Table 5 for comparison, including various techniques: graph-cut based [55], shape model based [56], multi-atlas based [57,58], self-supervised 3D feature learning [59].

The obtained results show that our proposed method successfully generalizes and performs well on previously unseen data too. Of course, the performance is slightly weaker when compared to methods trained and tested on the VISCERAL database itself. However, in the absence of training data, which is often the case in real life applications, the achieved performance is promising and the proposed approach can be applied efficiently.

4. Discussion

Quantitative experiments, applying 6 different evaluation metrics show that the proposed approach can detect the liver with high accuracy. For the majority of the metrics, the introduced method outperforms the compared state-of-the-art approaches. The overall performance, measured by the DSC is 0.964 for the MICCAI Sliver07 and 0.945 for the 3D-IRCADb database, which is the second best for MICCAI Sliver07 (Table 2) and third best achieved value for 3D-IRCADb (Table 3). However, our method achieves average VOE value and the second best RVD value, which should be further discussed. Concerning the fact, that for both the MICCAI Sliver07 and for the 3D-IRCADb database the RVD has negative value, we may conclude that the proposed method usually undersegment the liver organ, resulting in a relatively smaller output, than the ground truth. This undersegmentation can also explain the higher VOE value. Typical undersegmentation errors are also presented in Fig. 14. After overviewing the segmentation results, we present some problematic slices from the volume with the weakest performance from MICCAI Sliver07 databases to give a further analysis (Fig. 15). The original scanned slice, the preprocessed input and the final segmentation result (in yellow) together with the ground truth (in red) are also present. We have selected three different slices which introduce the main segmentation errors of the volume.

The selected volume in Fig. 15 has a large tumor on the liver border, which is not detected (first row of Fig. 15). However, the whole organ, including the tumor part correctly appears in the preprocessed input. In the second row, the undersegmentation of the tumor part still exists, moreover the segmentation is also inaccurate in the surroundings of the hepatic arteries. It should be mentioned, that one drawback of the preprocessing might be that intensity variation inside the liver is also amplified in the preprocessed input (center column of Fig. 15), which may complicates the accurate segmentation. In the bottom row of Fig. 15, the tumor part is already accurately segmented, probably because it is not located on the liver border, but inside the organ. However, the organ consists of two separate parts according to the annotated ground truth, which is missed by the proposed method and the two parts are merged during the segmentation. Analyzing the quantitative evaluation metrics, the RVD has a higher negative value of −11.166 caused by the missed tumor parts. Similarly, these larger segmentation errors result in a higher VOE value (14.239) and a reduced DSC value (0.923). Therefore, the method could be further improved for handling liver tumors more efficiently, which could also enhance the quantitative segmentation performance as well.



(a) Original scan (b) Preprocessed input (c) Result and GT

Fig. 15. Segmentation errors on selected slices of the MICCAI Sliver07 scan with the weakest quantitative result. DSC: 0.923, VOE: 14.239, RVD: −11.166. The contour of the ground truth (GT) is in red. Yellow contour denotes the result of proposed method.

The performance of the proposed preprocessing step is also tested, the learning-based segmentation is also evaluated with a simpler preprocessing, however, the majority of the quantitative values is significantly improved when applying the proposed preprocessing. This may demonstrate that the preprocessing also contributes to the proposed method's accuracy.

Beside the segmentation performance, it is also important to note, that our aim is to construct a method, which can be used in a real life application, therefore the computational time and the adaptability for heterogeneous data are also important. Our proposed method has three main steps: the preprocessing takes 8.4 seconds, the learning-based prediction takes 6.2 seconds and the postprocessing step is 1.2 seconds. Altogether, the complete process takes 15.8 seconds on average for a CT scan measured on the MICCAI Sliver07 database. The algorithm was implemented in Matlab (pre- and postprocessing steps) and Python (prediction step), and computational time was measured on a simple PC with Intel(R) Core(TM) i9-9900 CPU @3.60 GHz and 64 GB RAM. The proposed method is rather fast compared to the other state-of-the-art approaches, complex networks usually requires more time for segmentation: [31] takes $30-200$ seconds for one CT scan, depending on the number of slices; [34] runs for 68 s for a MICCAI Sliver07 volume on average. Shape-based methods usually need more time: [17] indicates an average of 750 s computational time for one volume, [18] needs $5-6$ seconds for just one slice, [20] requires more than 16 min to perform the segmentation on one volume, however [16] is quite fast and it takes 45 s on average to segment one CT scan.

On the other hand, the size of the neural network is an important issue, as the model complexity is restricted by the GPU memory. Our model has approximately $17,700,000$ parameters in total, which is a medium sized network. As for comparison, H-DenseUNet, introduced in [31], explores hybrid (intra-slice and inter-slice) features using a 2D and 3D DenseUNet structure, resulting in more than $200,000,000$ parameters, which requires huge GPU memory and longer computation time.

Concerning the adaptability of the proposed method and the performance on previously unseen data, the evaluation results are quite promising; the performance does not deteriorate drastically, however there is a slight relapse. This evaluation step validated that our method is not overtrained, but actually performs robustly in case of segmenting heterogeneous 3D medical data. In the future, the method should be tested on CT volumes from multiple data sources to further analyze the weaknesses and improve them, so that the approach could be integrated into a real life application.

5. Conclusion

In this paper, we presented an automatic, deep learning based approach, which is able to handle smaller databases including heterogeneous CT data. Our aim was to develop an automated

method, which integrates the advantages of pixel-based and learning-based approaches, thus performs with high accuracy and low computational time even for previously unseen data. The proposed method starts with a preprocessing, applying probability density function based estimation and supervoxel segmentation to remove most of the non-liver voxels of the CT scan. The preprocessed scan is then segmented using a proposed modification of the 3D-U-Net, which is called the 3D RP-UNet. As a novelty, the ResPath is introduced in the 3D Network, which is able to process the whole CT scan with high segmentation accuracy. To compensate for some typical segmentation errors (oversegmentation at the heart or spleen and undersegmentation caused by liver lesions), a postprocessing step is proposed, including a liver-heart separation and morphological refinement.

Quantitative evaluation is performed on three public databases (MICCAI Sliver07, 3D-IRCADb and VISCERAL) using 6 different evaluation metrics. The MICCAI Sliver07 and 3D-IRCADb databases are used for training and testing. The obtained results confirm that the proposed algorithm performs effectively and concerning the majority of the measured evaluation metrics, it achieves strong segmentation performance compared to the state-of-the-art approaches. Moreover, the trained 3D network was also tested for the previously unseen VISCERAL database, and the results showed that our method is capable of generalization and performs quite robustly on heterogeneous data with different image characteristics. Our future plan includes the improvement of the 3D CNN model to further integrate top-down information into the network model itself to make the performance more robust on heterogeneous data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The research was supported by the Ministry of Innovation and Technology NRDI Office within the framework of the Artificial Intelligence National Laboratory Program. Andrea Manno-Kovacs was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

[1] T. Heimann, B. Van Ginneken, M.A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, et al., Comparison and evaluation of methods for liver segmentation from CT datasets, IEEE T. Med. Imaging 28 (8) (2009) 1251–1265.
[2] E. van Rikxoort, Y. Arzhaeva, B. van Ginneken, Automatic segmentation of the liver in computed tomography scans with voxel classification and atlas matching, Proc. MICCAI Wshp. 3D Segmentation in the Clinic: A Grand Challenge (2007) 101–108.
[3] L. Rusko, G. Bekes, G. Nemeth, M. Fidrich, Fully automatic liver segmentation for contrast-enhanced CT images, Proc. MICCAI Wshp. 3D Segmentation in the Clinic: A Grand Challenge 2 (7).
[4] L. Huang, M. Weng, H. Shuai, Y. Huang, J. Sun, F. Gao, Automatic liver segmentation from CT images using single-block linear detection, BioMed. Res. Int. (2016).
[5] W. Wu, Z. Zhou, S. Wu, Y. Zhang, Automatic liver segmentation on volumetric CT images using supervoxel-based graph cuts, Comput. Math. Method M (2016).
[6] A. Zareei, A. Karimi, Liver segmentation with new supervised method to create initial curve for active contour, Comput. Biol. Med. 75 (2016) 139–150.
[7] N. Satpute, J. Gómez-Luna, J. Olivares, Accelerating Chan-Vese model with cross-modality guided contrast enhancement for liver segmentation, Comput. Biol. Med. 124 (2020) 103930.
[8] X. Yang, J.D. Yang, H.P. Hwang, H.C. Yu, S. Ahn, B.-W. Kim, H. You, Segmentation of liver and vessels from CT images and classification of liver segments for preoperative liver surgical planning in living donor liver transplantation, Comput. Meth. Prog. Bio. 158 (2018) 41–52.
[9] L. Xu, Y. Zhu, Y. Zhang, H. Yang, Liver segmentation based on region growing and level set active contour model with new signed pressure force function, Optik 202 (2020) 163705.
[10] Q. Huang, H. Ding, X. Wang, G. Wang, Fully automatic liver segmentation in CT images using modified graph cuts and feature detection, Comput. Biol. Med. 95 (2018) 198–208.
[11] T. Siriapisith, W. Kusakunniran, P. Haddawy, Pyramid graph cut: Integrating intensity and gradient information for grayscale medical image segmentation, Comput. Biol. Med. 126 (2020) 103997.
[12] R. Beichel, A. Bornik, C. Bauer, E. Sorantin, Liver segmentation in contrast enhanced CT data using graph cuts and interactive 3D segmentation refinement methods, Med. Phys. 39 (3) (2012) 1361–1373.
[13] A. Afifi, T. Nakaguchi, Liver segmentation approach using graph cuts and iteratively estimated shape and intensity constrains, in: Proc. Int. Conf. Medical Image Computing and Computer-assisted Intervention (MICCAI), 2012, pp. 395–403.
[14] M. Liao, Y.-Q. Zhao, X.-Y. Liu, Y.-Z. Zeng, B.-J. Zou, X.-F. Wang, F.Y. Shih, Automatic liver segmentation from abdominal CT volumes using graph cuts and border marching, Comput. Meth. Prog. Bio. 143 (2017) 1–12.
[15] X. Zhang, J. Tian, K. Deng, Y. Wu, X. Li, Automatic liver segmentation using a statistical shape model with optimal surface detection, IEEE T. Bio-med. Eng. 57 (10) (2010) 2622–2626.
[16] M. Erdt, S. Steger, M. Kirschner, S. Wesarg, Fast automatic liver segmentation combining learned shape priors with observed shape deviation, in: Proc. IEEE Int. Symp. Computer-Based Medical Systems (CBMS), IEEE, 2010, pp. 249–254.
[17] M. Esfandiarkhani, A.H. Foruzan, A generalized active shape model for segmentation of liver in low-contrast CT volumes, Comput. Biol. Med. 82 (2017) 59–70.
[18] M.-A. Lebre, A. Vacavant, M. Grand-Brochier, H. Rositi, R. Strand, H. Rosier, A. Abergel, P. Chabrot, B. Magnin, A robust multi-variability model based liver segmentation algorithm for CT-scan and MRI modalities, Comput. Med. Imag. Grap. 76 (2019) 101635.
[19] X. Lu, Q. Xie, Y. Zha, D. Wang, Fully automatic liver segmentation combining multi-dimensional graph cut with shape information in 3D CT images, Sci. Rep. 8 (1) (2018) 1–9.
[20] Y. Li, Y.-Q. Zhao, F. Zhang, M. Liao, L.-L. Yu, B.-F. Chen, Y.-J. Wang, Liver segmentation from abdominal CT volumes based on level set and sparse shape composition, Comput. Meth. Prog. Bio. 195 (2020) 105533.
[21] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Proc. Int. Conf. Medical Image Computing and Computer-assisted Intervention (MICCAI), 2015, pp. 234–241.
[22] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: Proc. Int. Conf. Medical Image Computing and Computer-assisted Intervention (MICCAI), Springer, 2016, pp. 424–432.
[23] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, P.-A. Heng, 3D deeply supervised network for automatic liver segmentation from CT volumes, in: Proc. Int. Conf. Medical Image Computing and Computer-assisted Intervention (MICCAI), Springer, 2016, pp. 149–157.
[24] P.F. Christ, M.E.A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields, in: Proc. Int. Conf. Medical Image Computing and Computer-assisted Intervention (MICCAI), Springer, 2016, pp. 415–423.
[25] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, L. Jin, X. Liu, X. Li, X. Qian, Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs, Artif. Intell. Med. 83 (2017) 58–66.
[26] G. Chlebus, A. Schenk, J.H. Moltz, B. van Ginneken, H.K. Hahn, H. Meine, Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing, Sci. Rep. 8 (1) (2018) 1–7.
[27] M. Ahmad, D. Ai, G. Xie, S.F. Qadri, H. Song, Y. Huang, Y. Wang, J. Yang, Deep belief network modeling for automatic liver segmentation, IEEE Access 7 (2019) 20585–20595.
[28] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, K.H. Maier-Hein, No new-net, in: Proc. Int. MICCAI Brainlesion Workshop, 2018, pp. 234–244.
[29] E. Caver, L. Chang, W. Zong, Z. Dai, N. Wen, Automatic brain tumor segmentation using a U-net neural network, in: Proc. Int. MICCAI BraTS Challenge, 2018, pp. 63–73.
[30] N. Ibtehaz, M.S. Rahman, MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation, Neural Networks 121 (2020) 74–87.
[31] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, P.-A. Heng, H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes, IEEE T. Med. Imaging 37 (12) (2018) 2663–2674.
[32] H. Seo, C. Huang, M. Bassenne, R. Xiao, L. Xing, Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images, IEEE T. Med. Imaging 39 (5) (2019) 1316–1325.
[33] Ü. Budak, Y. Guo, E. Tanyildizi, A. Şengür, Cascaded deep convolutional encoder-decoder neural networks for efficient liver tumor segmentation, Med. Hypotheses 134 (2020) 109431.
[34] J. Wang, P. Lv, H. Wang, C. Shi, SAR-U-Net: Squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in computed tomography, Comput. Meth. Prog. Bio. 208 (2021) 106268.

[35] W. Tang, D. Zou, S. Yang, J. Shi, J. Dan, G. Song, A two-stage approach for automatic liver segmentation with faster R-CNN and DeepLab, Neural Comput. Appl. 1–10 (2020).
[36] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE T. Pattern Anal. Mach. Intell. 39 (6) (2016) 1137–1149.
[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30.
[38] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, ICLR.
[39] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7794–7803.
[40] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, J. Shlens, Stand-alone self-attention in vision models, Advances in Neural Information Processing Systems 32.
[41] X. Yan, H. Tang, S. Sun, H. Ma, D. Kong, X. Xie, After-unet: Axial fusion transformer unet for medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3971–3981.
[42] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306.
[43] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, D. Xu, Unetr: Transformers for 3d medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 574–584.
[44] Y. Tang, D. Yang, W. Li, H. Roth, B. Landman, D. Xu, V. Nath, A. Hatamizadeh, Self-supervised pre-training of swin transformers for 3d medical image analysis, arXiv preprint arXiv:2111.14791.
[45] W. Xu, H. Liu, X. Wang, Y. Qian, Liver segmentation in CT based on ResUNet with 3D probabilistic and geometric post process, in: 2019 IEEE 4th Int. Conf. Signal and Image Processing (ICSIP), IEEE, 2019, pp. 685–689.
[46] L. Soler, A. Hostettler, V. Agnus, A. Charnoz, J.-B. Fasquel, J. Moreau, A.-B. Osswald, M. Bouhadjar, J. Marescaux, 3D image reconstruction for comparison of algorithm database, Tech. rep., IRCAD, Strasbourg, France, (accessed: 23.09.2019) (2010). https://www.ircad.fr/research/3dircadb/.
[47] G. Langs, A. Hanbury, B. Menze, H. Müller, VISCERAL: towards large data in medical imaging–challenges and directions, in: Proc. MICCAI Int. Wshp. Medical Content-based Retrieval for Clinical Decision Support, Springer, 2012, pp. 92–98.
[48] V. Czipczer, A. Manno-Kovacs, Automatic liver segmentation on CT images combining region-based techniques and convolutional features, in: Proc. Int. Conf. Content-Based Multimedia Indexing (CBMI), IEEE, 2019, pp. 1–6.
[49] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (11) (2012) 2274–2282.
[50] L. Yu, X. Yang, H. Chen, J. Qin, P.A. Heng, Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images, in: Thirty-first AAAI Conf. Artificial Intelligence, 2017.
[51] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, C. Pal, The importance of skip connections in biomedical image segmentation, in: Proc. Int. Wshp. Deep Learning and Data Labeling for Medical Applications, Springer, 2016, pp. 179–187.
[52] F. Lu, F. Wu, P. Hu, Z. Peng, D. Kong, Automatic 3D liver location and segmentation via convolutional neural network and graph cut, Int. J. Comput. Ass. Rad. 12 (2) (2017) 171–182.
[53] B. He, D. Yin, X. Chen, H. Luo, D. Xiao, M. He, G. Wang, C. Fang, L. Liu, F. Jia, A study of generalization and compatibility performance of 3d u-net segmentation on multiple heterogeneous liver ct datasets, BMC Medical Imaging 21 (1) (2021) 1–13.
[54] P. Bilic, P.F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser, et al., The liver tumor segmentation benchmark (lits), arXiv preprint arXiv:1901.04056.
[55] R. Kéchichian, S. Valette, M. Sdika, M. Desvignes, Automatic 3d multiorgan segmentation via clustering and graph cut using spatial relations and hierarchically-registered atlases, in: International MICCAI Workshop on Medical Computer Vision, Springer, 2014, pp. 201–209.
[56] B. He, C. Huang, F. Jia, Fully automatic multi-organ segmentation based on multi-boost learning and statistical shape model search, in: VISCERAL Challenge ISBI, 2015, pp. 18–21.
[57] T. Gass, G. Szekely, O. Goksel, Multi-atlas segmentation and landmark localization in images with large field of view, in: International MICCAI Workshop on Medical Computer Vision, Springer, 2014, pp. 171–180.
[58] B. Oliveira, S. Queirós, P. Morais, H.R. Torres, J. Gomes-Fonseca, J.C. Fonseca, J.L. Vilaça, A novel multi-atlas strategy with dense deformation field reconstruction for abdominal and thoracic multi-organ segmentation from computed tomography, Medical Image Analysis 45 (2018) 108–120.
[59] M. Blendowski, H. Nickisch, M.P. Heinrich, How to learn from unlabeled volume data: Self-supervised 3D context feature learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 649–657.

Vanda Czipczer is an info-bioincs engineering M.Sc. student at the Pazmany Peter Catholic University, Faculty of Information Technology and Bionics. She received her B.Sc. in Molecular bionics engineering. She works in a software developer position at the Machine Perception Research Laboratory at the Institute for Computer Science and Control (SZTAKI), Budapest, Hungary. Main research interest: medical applications in information technology.

Dr. Andrea Manno-Kovacs received her PhD in image processing in 2013 from the Pazmany Peter Catholic University in Budapest, her MSc in computer science from Budapest University of Technology and Economics. She is currently a senior research fellow with the Machine Perception Research Laboratory, at the Institute for Computer Science and Control (SZTAKI). Her research interests include image and video processing, feature extraction, saliency models and boundary recognition.