

Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields

Patrick Ferdinand Christ¹, Mohamed Ezzeldin A. Elshaer¹, Florian Ettl¹,
Sunil Tatavarthy², Marc Bickel¹, Patrick Bilic¹, Markus Rempfler¹, Marco
Armbruster⁴, Felix Hofmann⁴, Melvin D’Anastasi⁴, Wieland H. Sommer⁴
Seyed-Ahmad Ahmadi³, and Bjoern H. Menze¹

¹ Technische Universität München, Image-Based Biomedical Modeling Group,
Arcisstrasse 21, 80333 Munich, Germany

Patrick.Christ@tum.de and Bjoern.Menze@tum.de

² Technische Universität München, Chair for Data Processing, Arcisstrasse 21,
80333 Munich, Germany

³ LMU Hospital Grosshadern, Department for Neurology, Marchioninistrasse 15,
81377 Munich, Germany

⁴ LMU Hospital Grosshadern, Department for Clinical Radiology, Marchioninistrasse
15, 81377 Munich, Germany

Abstract. Automatic segmentation of the liver and its lesion is an important step towards deriving quantitative biomarkers for accurate clinical diagnosis and computer-aided decision support systems. This paper presents a method to automatically segment liver and lesions in CT abdomen images using cascaded fully convolutional neural networks (CFCNs) and dense 3D conditional random fields (CRFs). We train and cascade two FCNs for a combined segmentation of the liver and its lesions. In the first step, we train a FCN to segment the liver as ROI input for a second FCN. The second FCN solely segments lesions from the predicted liver ROIs of step 1. We refine the segmentations of the CFCN using a dense 3D CRF that accounts for both spatial coherence and appearance. CFCN models were trained in a 2-fold cross-validation on the abdominal CT dataset 3DIRCAD comprising 15 hepatic tumor volumes. Our results show that CFCN-based semantic liver and lesion segmentation achieves Dice scores over 94% for liver with computation times below 100s per volume. We experimentally demonstrate the robustness of the proposed method as a decision support system with a high accuracy and speed for usage in daily clinical routine.

Keywords: Liver, Lesion, Segmentation, FCN, CRF, CFCN, Deep Learning

1 Introduction

Anomalies in the shape and texture of the liver and visible lesions in CT are important biomarkers for disease progression in primary and secondary hepatic tumor disease [9]. In clinical routine, manual or semi-manual techniques

are applied. These, however, are subjective, operator-dependent and very time-consuming. In order to improve the productivity of radiologists, computer-aided methods have been developed in the past, but the challenges in automatic segmentation of combined liver and lesion remain, such as low-contrast between liver and lesion, different types of contrast levels (hyper-/hypo-intense tumors), abnormalities in tissues (metastases), size and varying amount of lesions.

Nevertheless, several interactive and automatic methods have been developed to segment the liver and liver lesions in CT volumes. In 2007 and 2008, two Grand Challenges benchmarks on liver and liver lesion segmentation have been conducted [9,4]. Methods presented at the challenges were mostly based on statistical shape models. Furthermore, grey level and texture based methods have been developed [9]. Recent work on liver and lesion segmentation employs graph cut and level set techniques [16,15,17], sigmoid edge modeling [5] or manifold and machine learning [11,6]. However, these methods are not widely applied in clinics, due to their speed and robustness on heterogeneous, low-contrast real-life CT data. Hence, interactive methods were still developed [7,1] to overcome these weaknesses, which yet involve user interaction.

Deep Convolutional Neural Networks CNN have gained new attention in the scientific community for solving computer vision tasks such as object recognition, classification and segmentation [14,18], often out-competing state-of-the-art methods. Most importantly, CNN methods have proven to be highly robust to varying image appearance, which motivates us to apply them to fully automatic liver and lesions segmentation in CT volumes.

Semantic image segmentation methods based on fully convolutional neural networks FCN were developed in [18], with impressive results in natural image segmentation competitions [3,24]. Likewise, new segmentation methods based on CNN and FCNs were developed for medical image analysis, with highly competitive results compared to state-of-the-art. [20,8,23,21,19,12].

In this work, we demonstrate the combined automatic segmentation of the liver and its lesions in low-contrast heterogeneous CT volumes. Our contributions are three-fold. First, we train and apply fully convolutional CNN on CT volumes of the liver for the first time, demonstrating the adaptability to challenging segmentation of hepatic liver lesions. Second, we propose to use a cascaded fully convolutional neural network (CFCN) on CT slices, which segments liver and lesions sequentially, leading to significantly higher segmentation quality. Third, we propose to combine the cascaded CNN in 2D with a 3D dense conditional random field approach (3DCRF) as a post-processing step, to achieve higher segmentation accuracy while preserving low computational cost and memory consumption. In the following sections, we will describe our proposed pipeline (Section 2.2) including CFCN (Section 2.3) and 3D CRF (Section 2.4), illustrate experiments on the 3DIRCADb dataset (Section 2) and summarize the results (Section 4).

2 Methods

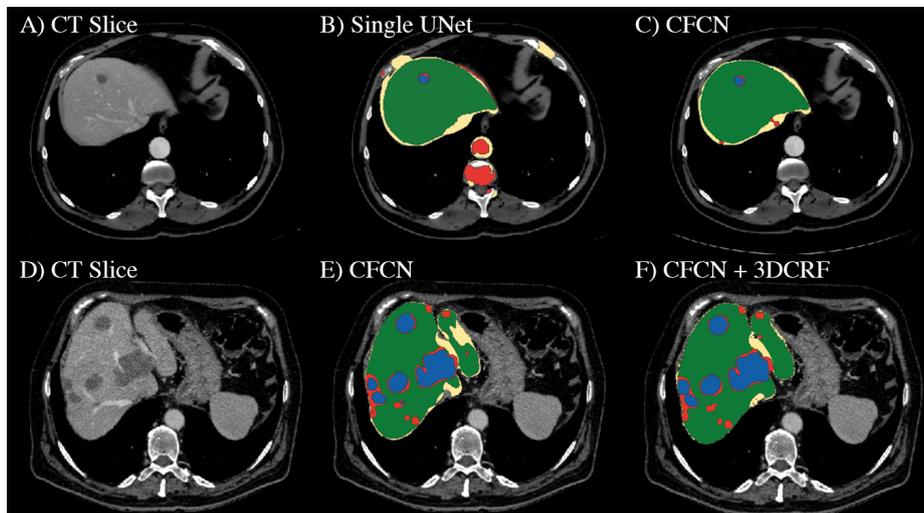


Fig. 1. Automatic liver and lesion segmentation with cascaded fully convolutional networks (CFCN) and dense conditional random fields (CRF). Green depicts correctly predicted liver segmentation, yellow for liver false negative and false positive pixels (all wrong predictions), blue shows correctly predicted lesion segmentation and red lesion false negative and false positive pixels (all wrong predictions). In the first row, the false positive lesion prediction in B of a single UNet as proposed by [20] were eliminated in C by CFCN as a result of restricting lesion segmentation to the liver ROI region. In the second row, applying the 3DCRF to CFCN in F increases both liver and lesion segmentation accuracy further, resulting in a lesion Dice score of 82.3%.

In the following section, we denote the 3D image volume as I , the total number of voxels as N and the set of possible labels as $\mathcal{L} = \{0, 1, \dots, l\}$. For each voxel i , we define a variable $x_i \in \mathcal{L}$ that denotes the assigned label. The probability of a voxel i belonging to label k given the image I is described by $P(x_i = k|I)$ and will be modelled by the FCN. In our particular study, we use $\mathcal{L} = \{0, 1, 2\}$ for background, liver and lesion, respectively.

2.1 3DIRCADb Dataset

For clinical routine usage, methods and algorithms have to be developed, trained and evaluated on heterogeneous real-life data. Therefore, we evaluated our proposed method on the 3DIRCADb dataset⁵[22]. In comparison to the the grand

⁵ The dataset is available on <http://ircad.fr/research/3d-ircadb-01>

challenge datasets, the 3DIRCADb dataset offers a higher variety and complexity of livers and its lesions and is publicly available. The 3DIRCADb dataset includes 20 venous phase enhanced CT volumes from various European hospitals with different CT scanners. For our study, we trained and evaluated our models using the 15 volumes containing hepatic tumors in the liver with 2-fold cross validation. The analyzed CT volumes differ substantially in the level of contrast-enhancement, size and number of tumor lesions (1 to 42). We assessed the performance of our proposed method using the quality metrics introduced in the grand challenges for liver and lesion segmentation by [9,4].

2.2 Data preparation, processing and pipeline

Pre-processing was carried out in a slice-wise fashion. First, the Hounsfield unit values were windowed in the range $[-100, 400]$ to exclude irrelevant organs and objects, then we increased contrast through histogram equalization. As in [20], to teach the network the desired invariance properties, we augmented the data by applying translation, rotation and addition of gaussian noise. Thereby resulting in an increased training dataset of 22,693 image slices, which were used to train two cascaded FCNs based on the UNet architecture [20]. The predicted segmentations are then refined using dense 3D Conditional Random Fields. The entire pipeline is depicted in Figure 2.

2.3 Cascaded Fully Convolutional Neural Networks (CFCN)

We used the UNet architecture [20] to compute the soft label probability maps $P(x_i|I)$. The UNet architecture enables accurate pixel-wise prediction by combining spatial and contextual information in a network architecture comprising 19 convolutional layers. In our method, we trained one network to segment the liver in abdomen slices (step 1), and another network to segment the lesions, given an image of the liver (step 2). The segmented liver from step 1 is cropped and resampled to the required input size for the cascaded UNet in step 2, which further segments the lesions.

The motivation behind the cascade approach is that it has been shown that UNets and other forms of CNNs learn a hierarchical representation of the provided data. The stacked layers of convolutional filters are tailored towards the desired classification in a data-driven manner, as opposed to designing hand-crafted features for separation of different tissue types. By cascading two UNets, we ensure that the UNet in step 1 learns filters that are specific for the detection and segmentation of the liver from an overall abdominal CT scan, while the UNet in step 2 arranges a set of filters for separation of lesions from the liver tissue. Furthermore, the liver ROI helps in reducing false positives for lesions.

A crucial step in training FCNs is appropriate class balancing according to the pixel-wise frequency of each class in the data. In contrast to [18], we observed that training the network to segment small structures such as lesions is not possible without class balancing, due to the high class imbalance. Therefore we introduced an additional weighting factor ω^{class} in the cross entropy loss

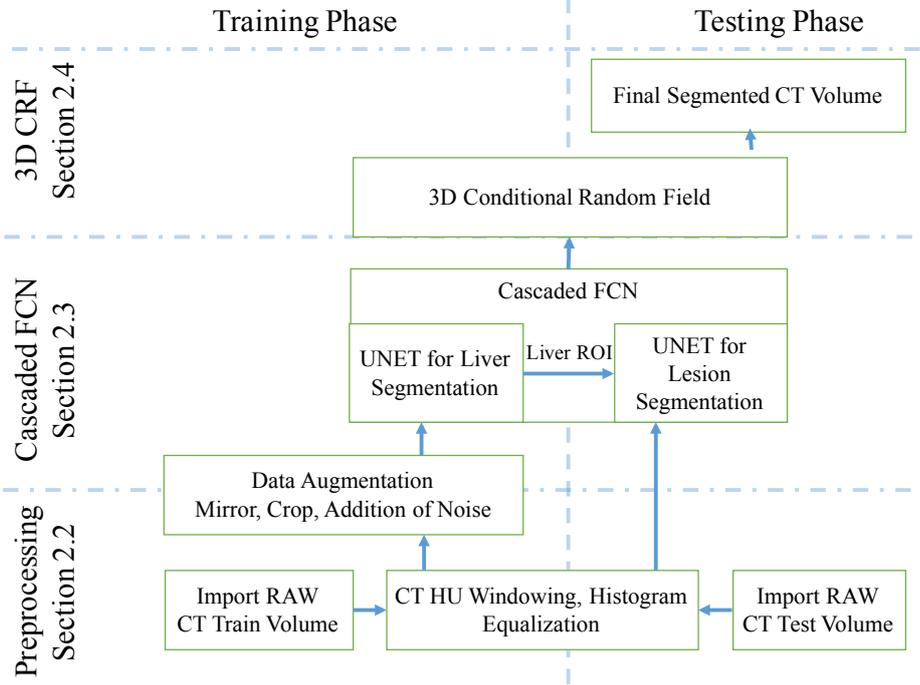


Fig. 2. Overview of the proposed image segmentation pipeline. In the training phase, the CT volumes are trained after pre-processing and data augmentation in a cascaded fully convolutional neural network (CFCN). To gain the final segmented volume, the test volume is fed-forward in the (CFCN) and refined afterwards using a 3D conditional random field 3DCRF.

function L of the FCN.

$$L = -\frac{1}{n} \sum_{i=1}^N \omega_i^{class} \left[\hat{P}_i \log P_i + (1 - \hat{P}_i) \log(1 - P_i) \right] \quad (1)$$

P_i denotes the probability of voxel i belonging to the foreground, \hat{P}_i represents the ground truth. We chose ω_i^{class} to be $\frac{1}{|\text{Pixels of Class } x_i=k|}$.

The CFCNs were trained on a NVIDIA Titan X GPU, using the deep learning framework caffe [10], at a learning rate of 0.001, a momentum of 0.8 and a weight decay of 0.0005.

2.4 3D Conditional Random Field (3DCRF)

Volumetric FCN implementation with 3D convolutions is strongly limited by GPU hardware and available VRAM [19]. In addition, the anisotropic resolution of medical volumes (e.g. 0.57-0.8mm in xy and 1.25-4mm in z voxel dimension in 3DIRCADb) complicates the training of discriminative 3D filters. Instead, to

capitalise on the locality information across slices within the dataset, we utilize 3D dense conditional random fields CRFs as proposed by [13]. To account for 3D information, we consider all slice-wise predictions of the FCN together in the CRF applied to the entire volume at once.

We formulate the final label assignment given the soft predictions (probability maps) from the FCN as *maximum a posteriori* (MAP) inference in a dense CRF, allowing us to consider both spatial coherence and appearance.

We specify the dense CRF following [13] on the complete graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $i \in \mathcal{V}$ for each voxel in the image and edges $e_{ij} \in \mathcal{E} = \{(i, j) \mid \forall i, j \in \mathcal{V} \text{ s.t. } i < j\}$ between *all* vertices. The variable vector $\mathbf{x} \in \mathcal{L}^N$ describes the label of each vertex $i \in \mathcal{V}$. The energy function that induces the according Gibbs distribution is then given as:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \phi_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \phi_{ij}(x_i, x_j) , \quad (2)$$

where $\phi_i(x_i) = -\log P(x_i|I)$ are the unary potentials that are derived from the FCNs probabilistic output, $P(x_i|I)$. $\phi_{ij}(x_i, x_j)$ are the pairwise potentials, which we set to:

$$\begin{aligned} \phi_{ij}(x_i, x_j) = & \mu(x_i, x_j) \left(w_{\text{pos}} \exp \left(-\frac{|p_i - p_j|^2}{2\sigma_{\text{pos}}^2} \right) \right. \\ & \left. + w_{\text{bil}} \exp \left(-\frac{|p_i - p_j|^2}{2\sigma_{\text{bil}}^2} - \frac{|I_i - I_j|^2}{2\sigma_{\text{int}}^2} \right) \right) , \end{aligned} \quad (3)$$

where $\mu(x_i, x_j) = \mathbf{1}(x_i \neq x_j)$ is the Potts function, $|p_i - p_j|$ is the spatial distance between voxels i and j and $|I_i - I_j|$ is their intensity difference in the original image. The influence of the pairwise terms can be adjusted with their weights w_{pos} and w_{bil} and their effective range is tuned with the kernel widths σ_{pos} , σ_{bil} and σ_{int} .

We estimate the best labelling $\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathcal{L}^N} E(\mathbf{x})$ using the efficient mean field approximation algorithm of [13]. The weights and kernels of the CRF were chosen using a random search algorithm.

3 Results and Discussion

The qualitative results of the automatic segmentation are presented in Figure 1. The complex and heterogeneous structure of the liver and all lesions were detected in the shown images. The cascaded FCN approach yielded an enhancement for lesions with respect to segmentation accuracy compared to a single FCN as can be seen in Figure 1. In general, we observe significant⁶ additional improvements for slice-wise Dice overlaps of liver segmentations, from mean Dice 93.1% to 94.3% after applying the 3D dense CRF.

Quantitative results of the proposed method are reported in Table 1. The CFCN achieves higher scores as the single FCN architecture. Applying the 3D

⁶ Two-sided paired t-test with p-value $< 4 \cdot 10^{-19}$

Approach	VOE [%]	RVD [%]	ASD [mm]	MSD [mm]	DICE [%]
UNET as in [20]	39	87	19.4	119	72.9
Cascaded UNET	12.8	-3.3	2.3	46.7	93.1
Cascaded UNET + 3D CRF	10.7	-1.4	1.5	24.0	94.3
Li et al. [16] (liver-only)	9.2	-11.2	1.6	28.2	
Chartrand et al. [2] (semi-automatic)	6.8	1.7	1.6	24	
Li et al. [15] (liver-only)					94.5

Table 1. Quantitative segmentation results of the liver on the 3DIRCADb dataset. Scores are reported as presented in the original papers.

CRF improved the segmentations results of calculated metrics further. The runtime per slice in the CFCN is $2 \cdot 0.2s = 0.4s$ without and $0.8s$ with CRF.

In comparison to state-of-the-art, such as [5,15,16,2], we presented a framework, which is capable of a combined segmentation of the liver and its lesion.

4 Conclusion

Cascaded FCNs and dense 3D CRFs trained on CT volumes are suitable for automatic localization and combined volumetric segmentation of the liver and its lesions. Our proposed method competes with state-of-the-art. We provide our trained models under open-source license allowing fine-tuning for other medical applications in CT data ⁷. Additionally, we introduced and evaluated dense 3D CRF as a post-processing step for deep learning-based medical image analysis. Furthermore, and in contrast to prior work such as [5,15,16], our proposed method could be generalized to segment multiple organs in medical data using multiple cascaded FCNs. All in all, heterogeneous CT volumes from different scanners and protocols as present in the 3DIRCADb dataset and in clinical trials can be segmented in under 100s each with the proposed approach. We conclude that CFCNs and dense 3D CRFs are promising tools for automatic analysis of liver and its lesions in clinical routine.

References

1. Ben-Cohen, A., et al.: Automated method for detection and segmentation of liver metastatic lesions in follow-up ct examinations. *Journal of Medical Imaging* (3) (2015)
2. Chartrand, G., et al.: Semi-automated liver ct segmentation using laplacian meshes. In: *ISBI*. pp. 641–644. *IEEE* (2014)
3. Chen, L.C., et al.: Semantic image segmentation with deep convolutional nets and fully connected crfs. *ICLR* (2015)

⁷ Trained models are available at <https://github.com/IBBM/Cascaded-FCN>

4. Deng, X., Du, G.: Editorial: 3d segmentation in the clinic: a grand challenge ii-liver tumor segmentation. In: MICCAI Workshop (2008)
5. Foruzan, A.H., Chen, Y.W.: Improved segmentation of low-contrast lesions using sigmoid edge model. *Int J Comput Assist Radiol Surg* pp. 1–17 (2015)
6. Freiman, M., Cooper, O., Lischinski, D., Joskowicz, L.: Liver tumors segmentation from cta images using voxels classification and affinity constraint propagation. *Int J Comput Assist Radiol Surg* 6(2), 247–255 (2011)
7. Häme, Y., Pollari, M.: Semi-automatic liver tumor segmentation with hidden markov measure field model and non-parametric distribution estimation. *Med Image Anal* 16(1), 140–149 (2012)
8. Havaei, M., et al.: Brain Tumor Segmentation with Deep Neural Networks. *Med Image Anal* (2016)
9. Heimann, T., et al.: Comparison and evaluation of methods for liver segmentation from ct datasets. *IEEE Trans. Med. Imag.* 28(8), 1251–1265 (Aug 2009)
10. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: *Proc ACM Int Conf Multimed.* pp. 675–678. ACM (2014)
11. Kadoury, S., Vorontsov, E., Tang, A.: Metastatic liver tumour segmentation from discriminant grassmannian manifolds. *Phys Med Biol* 60(16), 6459 (2015)
12. Kamnitsas, K., et. al.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *arXiv preprint arXiv:1603.05959* (2016)
13. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. In: *NIPS.* pp. 109–117 (2011)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *NIPS.* pp. 1097–1105 (2012)
15. Li, C., Wang, X., Eberl, S., Fulham, M., Yin, Y., Chen, J., Feng, D.D.: A likelihood and local constraint level set model for liver tumor segmentation from ct volumes. *IEEE Trans. Biomed. Eng.* 60(10), 2967–2977 (2013)
16. Li, G., Chen, X., Shi, F., Zhu, W., Tian, J., Xiang, D.: Automatic liver segmentation based on shape constraints and deformable graph cut in ct images. *IEEE Trans. Image Process.* 24(12), 5315–5329 (2015)
17. Linguraru, M.G., Richbourg, W.J., Liu, J., Watt, J.M., Pamulapati, V., Wang, S., Summers, R.M.: Tumor burden analysis on computed tomography by automated liver and tumor segmentation. *IEEE Trans. Med. Imag.* 31(10), 1965–1976 (2012)
18. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. *CVPR* (2015)
19. Prason, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: *MICCAI.* vol. 16, pp. 246–253 (2013)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *MICCAI,* vol. 9351, pp. 234–241 (2015)
21. Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B., Summers, R.M.: Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: *MICCAI,* pp. 556–564 (2015)
22. Soler, L., et al.: 3d image reconstruction for comparison of algorithm database: a patient-specific anatomical and medical image database (2012)
23. Wang, J., MacKenzie, J.D., Ramachandran, R., Chen, D.Z.: Detection of glands and villi by collaboration of domain knowledge and deep learning. In: *MICCAI,* pp. 20–27 (2015)
24. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.: Conditional random fields as recurrent neural networks. *ICCV* (2015)