Automatic multi-object organ detection and segmentation in abdominal CT-data

Oliver Mietzner [1] Andre Mastmeyer [2]

[1] Medical Informatics, Universität zu Lübeck, oliver.mietzner@student.uni-luebeck.de
[2] Digital Health Management, Hochschule Aalen, andre.mastmeyer@hs-aalen.de

Abstract

The ability to generate 3D patient models in a fast and reliable way, is of great importance, e.g. for the simulation of liver punctures in a virtual reality simulation [1], [2], [3], [4]. The aim is to automatically detect and segment abdominal structures in CT-scans. In particular among the selected organ group, the pancreas poses a challenge. We use a combination of random regression forests and U-Nets to detect bounding boxes and generate segmentation masks for five abdominal organs (liver, kidneys, spleen, pancreas). Training and testing is carried out on 50 CT-scans from various public sources. The results show Dice coefficients of up to 0.71.

1 Introduction

Virtual Reality (VR) intervention simulation for training and planning is slowly gaining importance in clinical teaching and routine. VR methods can be used for various tasks, ranging from training scenarios for the medical student and staff, to individual patient related simulations of planned operations. The necessary individual patient models should be available fast and be accurate to guarantee a plausible simulation.

The first step in producing patient models is to acquire patient image data. This data will be used for the following steps in the modelling process. It is desirable to use high quality image data, because inaccuracies will be carried over to the resulting model. The next step is the coarse localisation of the organs to be modelled inside organ-specific volumes of interest (VOI) to simplify their subsequent segmentation.

The aim of our study, is to automatically detect bounding boxes for organs in abdominal CT-data by a learning-from-example method. The detected bounding boxes are then used for segmentation. Detection of the abdominal organs can be a challenge, because of the variety of shapes these organs can have. The intensity based features of these organs are also challenging, because the intensity range of neighbouring organs often overlaps. We detected five abdominal organs (liver, right kidney, left kidney, pancreas, spleen), that are commonly used in simulations.

We used a Random-Forest method, that is developed by Criminisi et al. [5] to automatically detect the bounding boxes of abdominal organs. The method uses Random-Regression-Forests to predict the location of organ bounding boxes in CT-data. Furthermore, we used the U-Net, proposed by Ronneberger et al. [6], to automatically segment organs. The training and testing is carried out on a database of 50 abdominal CT-scans, that had to be segmented beforehand.

2 Material and Methods

The database, that is used for training and testing, consisted of 50 CT-scans from three different sources: SLIVER07[1], LiTS[2] and VISCERAL[3]. The data sets contained only abdominal CT-data. The sets also varied in quality and field of view, to capture the variability of organs and scans. The scans are not only different in the number of slices (64-861), thickness of slices (1-5mm) and Field-of-View (FoV), but also in image noise. This variability in the datasets is important, to ensure robustness vs. typical inference factors of clinical image data during training and application. All five target organs (liver, pancreas, left kidney, right kidney, spleen) are included in the scans.

Only a small fraction of the segmentation maps contained all five target organs, thus manual structure segmentation was necessary frequently. The segmentation of the target structures is carried out by using ITKSnap [4].

[1] http://sliver07.org
[2] http://competitions.codalab.org/competitions/17094
[3] http://visceral.eu/closed-benchmarks/anatomy3
[4] http://www.itksnap.org/pmwiki/pmwiki.php

Pre-Print-Manuscript
Studierendentagung Uni Lübeck 2020

2.1 Definition of ground truth bounding boxes

The segmentation maps associated to the CT-scans are the basis for the extraction of ground truth bounding boxes to be learned. A bounding box can be described as a cubical polyhedron that completely encloses an object. The form and size of a complex object is described by a much simpler hull object.

The corresponding organ bounding boxes of each of the scans can be created by scanning the segmentation maps for coordinate direction extremes of label occurrence. To create a three-dimensional bounding box, for each coordinate direction (x,y,z), we iterate through the segmentation map and save the maximum and minimum limits where a label is encountered.

2.2 Training of the models

Our method is composed of two different machine learning approaches, random regression forests (RRF) and U-Nets. The RRF's are used to detect organ bounding boxes in CT-data. The U-Nets make use of the detected bounding boxes and segment the organ contained in them.

2.2.1 Training random regression forests for bounding box detection

We use random regression forests to determine the location and extent of abdominal organs [5]. The main difference between random classification forests and RRFs, is the type of output that is predicted. While classification forests try to categorize objects, regression forests predict continuous values. Regression forests partition the data into manageable chunks to predict average values. As seen in Fig. 1, the method expects CT-scans and ground truth bounding boxes as input.

A three-dimensional bounding box b_c of an organ c can be described by using a six-dimensional vector $b_c = (b_c^L, b_c^R, b_c^A, b_c^P, b_c^H, b_c^F)$ with coordinates in mm [5][5]. We run over all voxels $p = (x_p, y_p, z_p)$, which are within a specified radial distance (r = 5cm) from the scan medial axis. The distance d between such a voxel and each of the bounding box walls, can be calculated by using $d(p) = (x_p - x^L, x_p - x^R, y_p - y^A, y_p - y^P, z_p - z^H, z_p - z^F)$ and is saved as the offset-vector to be learned. In contrast to Criminisi et al. [5], we use only 50 feature boxes, that are evenly distributed on three spherical shells (r = 5 cm; 2.5 cm; 1.25 cm) to generate the input feature vector. These feature boxes F_j are intended to capture the spatial and intensity context of the current voxel. For this purpose, the mean intensities of the feature boxes are calculated and saved in the feature vector. An example feature box is shown in Fig. 2. While training, the RRF learns the distance vector (later: output) to the reference bounding box using the feature boxes.

[5] L=left, R=right, A=anterior, P=posterior, H=head, F=foot

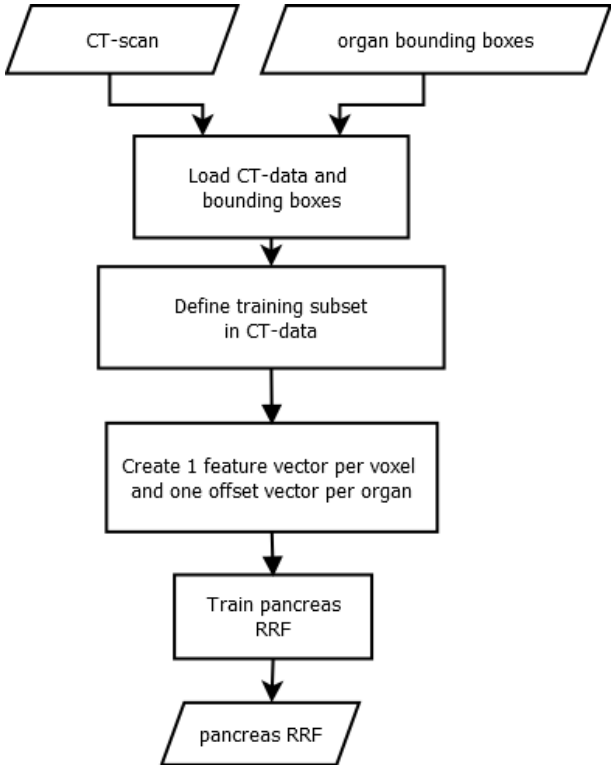


Figure 1: Training of a RRF: The inputs for the training process are CT-scans and ground truth bounding boxes of the targeted organs. We create one feature vector and one offset vector for each voxel that is part of a predefined medial cylinder subset in the scan [5]. The trained RRF is able to predict the offset between a voxel and an organ bounding box.

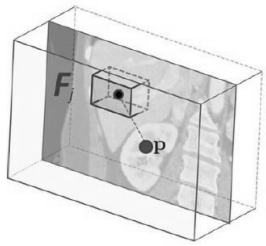


Figure 2: Example feature box: The feature box F_j is generated in correlation to the current voxel and calculated the mean value of a 3D image section [5].

2.2.2 Training of a U-Net for semantic segmentation

The training data for our U-Net consists of the expert segmentations and ground truth bounding boxes, as schematised in Fig. 3. The bounding boxes (VOIs) are then used to locally extract the intensity and label data from the CT-scans and their corresponding label maps. As input, the U-Net receives a VOI from the intensity data, while the same

Pre-Print-Manuscript
Studierendentagung Uni Lübeck 2020

region within the label data is connected to the output. We use the U-Net architecture proposed by Ronneberger et al. [6], which consists of nine layers with four down- and up-scale steps. We trained one U-Net for each organ.

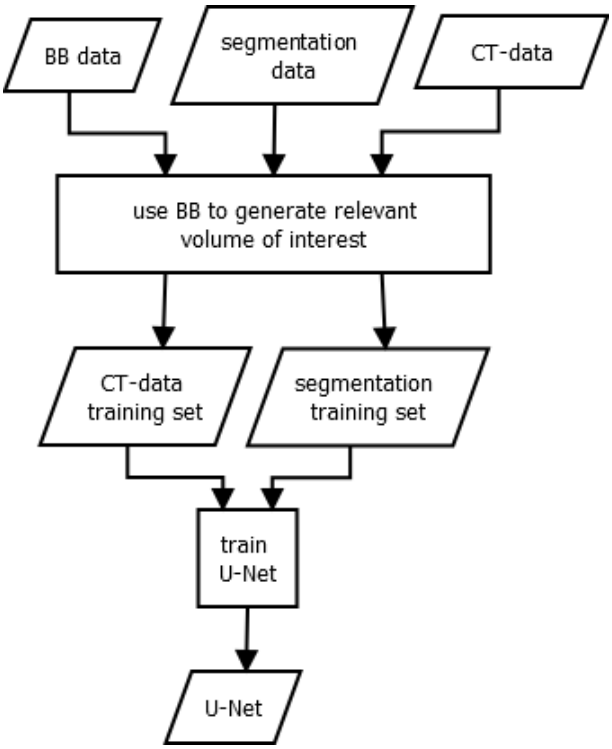


Figure 3: Training of a U-Net: The inputs for the training process are ground truth bounding boxes (BB), CT-scans and their corresponding segmentation maps. The box is applied to the CT- and segmentation data to extract the relevant image region. Inside the organ VOIs, the segmentation is learned. The process results in an organ-wise U-Net, that can segment image regions.

2.3 Application of the models

In the first step, the organ specific RRFs predict the organ bounding box candidates. Then, a distance vector is selected by majority voting and converted into a six-dimensional vector to describe the final organ-specific bounding box.

Now, the U-Net model gets the corresponding bounding box as input. The U-Net uses the data contained inside the given bounding box, to segment the corresponding organ. The output is a segmentation map of the full target organ.

2.4 Evaluation

We used a five-fold Monte Carlo cross-validation based on a 30:20 (train:test) data split. The resulting segmentations are compared against the reference segmentations by using the Dice coefficient [7].

3 Results and Discussion

Table 1: Comparison of mean Dice coefficients and standard deviation (SD) of our ensemble method with other one organ-focused methods.

Organ	Dice $\pm$ SD	Dice (of others)
liver	0.71$\pm$0.31	0.97 [8]
kidney (right)	0.55$\pm$0.34	0.91 [9]
kidney (left)	0.67$\pm$0.32	0.91 [9]
spleen	0.71$\pm$0.26	0.94 [10]
pancreas	0.32$\pm$0.28	0.74 [10]

Table 1 shows the Dice coefficients achieved by our method for all five target organs. We compared our values to studies, that tried to segment the same organs with automatic methods.

The best results show up for the liver and spleen. Both organs achieve a dice coefficient of 0.71. The segmentation of the pancreas are worst with only 0.32. The standard deviation is similar for all organs ranging from 0.26 to 0.34. An example segmentation is presented in Fig. 4.

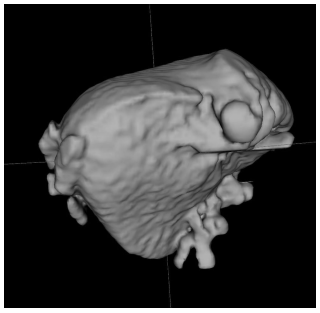
Segmentation of unseen patient image data scan takes between 30 seconds and one minute depending on the size of the data, i.e. on an Intel-i7 processor with a NVIDIA GTX 1050 GPU.

4 Conclusion

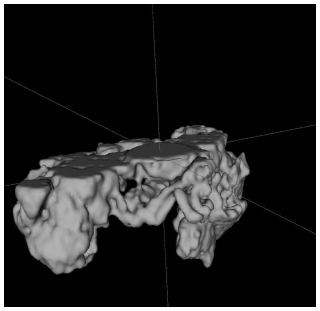
We were able to detect organ bounding boxes automatically. Nonetheless, the overall position and extent of the predicted bounding boxes has to be improved to ensure a satisfying segmentation. Especially the prediction along the z-axis is not accurate enough and resulted in boxes that are displaced. Due to misalignment, some segmentation masks are cut off, because parts of the organ are missing.The standard deviations are too high, indicating highly varying results.

In summary, the results of the current state of the segmentation method are affected by displaced bounding boxes and could not compete with the Dice coefficients of related studies. Though it is important to mention that most of the studies had more training data and are often focused on a single organ not addressing a group.

Future works should focus on the accuracy and robustness of the bounding box detection. The quality of the target segmentation depends heavily on the quality of the detected bounding boxes. We used predicted bounding boxes, to produce results for a completely automatic method.



(a)



(b)

Figure 4: Example segmentations: (a) A good segmentation of the liver, with some leakage into surrounding structures. (b) A poor segmentation of the pancreas. The bounding box was shifted along the z-axis. This cut off the upper half of the pancreas, resulting in an incomplete segmentation. The segmentation also included parts of surrounding structures

Acknowledgement

This work has been carried out with funding from DFG-MA 6791/1-1, Nvidia GPU grant and EXPLOR program by Stiftung Kessler + Co. für Bildung und Kultur.

5 References

[1] A. Mastmeyer, D. Fortmeier, and H. Handels, "Random forest classification of large volume structures for visuo-haptic rendering in ct images." p. 97842H, 2016.

[2] A. Mastmeyer, M. Wilms, D. Fortmeier, J. Schröder, and H. Handels, "Real-time ultrasound simulation for training of us-guided needle insertion in breathing virtual patients," in *Studies in health technology and informatics*, vol. 220. IOS Press, 2016, p. 219.

[3] A. Mastmeyer, M. Wilms, and H. Handels, "Interpatient respiratory motion model transfer for virtual reality simulations of liver punctures," *Journal of World Society of Computer Graphics - WSCG*, vol. 25, no. 1, pp. 1–10, 2017.

[4] ——, "Population-based respiratory 4d motion atlas construction and its application for vr simulations of liver punctures," in *SPIE Medical Imaging 2018: Image Processing*, vol. 10574. International Society for Optics and Photonics, 2018, p. 1057417.

[5] A. Criminisi, D. Robertson, O. Pauly, B. Glocker, E. Konukoglu, J. Shotton, D. Mateus, A. M. Möller, S. Nekolla, and N. Navab, *Decision Forests for Computer Vision and Medical Image Analysis*. London: Springer, 2013.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Med Image Comput Comput Assist Interv – MICCAI 2015*, p. 234–241, 2015.

[7] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC med imaging*, vol. 15, no. 1, p. 29, 2015.

[8] H. Meine, G. Chlebus, M. Ghafoorian, I. Endo, and A. Schenk, "Comparison of u-net-based convolutional neural networks for liver segmentation in CT," *CoRR*, vol. abs/1810.04017, 2018. [Online]. Available: http://arxiv.org/abs/1810.04017

[9] F. Isensee and K. H. Maier-Hein, "An attempt at beating the 3d u-net," 2019.

[10] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, "Automatic multi-organ segmentation on abdominal ct with dense v-networks," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1822–1834, Aug 2018.