2D vs. 3D U-Net Abdominal Organ Segmentation in CT Data using Organ Bounds

Daria Kern[a], Ulrich Klauck[a], Timo Ropinski[b], and Andre Mastmeyer[a]

[a]Aalen University, Faculties of Optics & Mechatronics, Electrical Engineering & Informatics, Aalen, Germany
[b]Ulm University, Institute of Media Informatics, Ulm, Germany

ABSTRACT

We compare axial 2D U-Nets and their 3D counterparts for voxel-based segmentation of five abdominal organs in CT scans. For each organ, two competing models are trained. The models are evaluated by performing five-fold cross-validation on 80 3D images. In a two-step concept, the relevant area containing the organ is first extracted using the ground truth bounding boxes and then passed as input to the U-Net. Furthermore, a random regression forest approach for the automatic generation of bounding boxes is summarized. The results show that the 2D U-Net is mostly on par with the 3D U-Net or even outperforms it. Especially for the kidneys, it is significantly better suited.

Keywords: 2D, 3D, U-Net, Architecture Comparison, Abdominal Organ Segmentation, CT Data, Random Regression Forest, Organ Bounding Box Detection

1. INTRODUCTION

The automatic reconstruction of 3D organ models from patient data has long been an unsolved challenge in medical image processing. Accurate and precise segmentation is especially important for VR-based intervention training and planning, where dynamic 4D Virtual Reality simulations are used.[1–3] Abdominal organs can be difficult to segment with regards to patient-individual variations, imaging noise, and low or varying organ contrast due to the contrast agent phase. The efficient creation of 3D models is still a bottleneck in research and the state-of-the-art. An automated generation of realistic 3D models under the aspect of saving computing power and memory with hardware restrictions is desirable. The saving of computing time and storage space is a highly relevant topic despite evolving potent hardware: For instance, mind big-data studies where many datasets are to be segmented in parallel on a limited number of computing nodes for the highest possible segmentation throughput or a continuous learning approach.

Among the various Deep Learning architectures used for semantic segmentation, the 2D U-Net[4] has proven particularly promising. A 3D version of the U-Net[5] was introduced in 2016. While a 3D model considers all three dimensions, a 2D model can only process two-dimensional data. Therefore, the data has to be examined slice-wise (i.e. sagittal, coronal, or axial). Splitting a 3D image into 2D slices results in more training images for a 2D model. In contrast to 2D, the processing of 3D data with a 3D model leads to an increased computing and memory effort in training and application.

H. Meine et al.[6] compare 2D and 3D U-Net architectures for liver segmentation in CT data. This preprint reports liver DSC above 0.97 using a 2.5D-technique and concludes that a 2D approach might be preferable to a 3D-approach. The 2.5D technique considers a fusion of results of all three orthogonal image planes (i.e. sagittal, coronal, axial). N. Takafumi et al.[7] focus on semantic segmentation of the lungs and compare 2D and 3D U-Nets with slightly different architectures. The results show no difference for mean DSC between the networks. However, the optimum learning rate was found to be 0.001 for the 2D U-Net and 0.0001 for the 3D U-Net.[7]

Further author information: (Send correspondence to Andre Mastmeyer)
Daria Kern: E-mail: hello@dariakern.de, Telephone: (+49) 7361 576 4567
Andre Mastmeyer: E-mail: andre.mastmeyer@hs-aalen.de, Telephone: (+49) 7361 576 4567

In this work, we investigate whether a 3D or an axial 2D architecture of the U-Net is better suited for the segmentation of five selected abdominal organs in volumetric CT scans. The studied organs are the liver, right kidney, left kidney, spleen, and pancreas. Since processing the entire 3D volume is computationally very expensive and negatively affects the quality of the segmentation, we first extract the relevant volume of interest (VOI) for each organ. Irrelevant structures are thus excluded and the segmentation of the relevant organs becomes easier and computationally more efficient. The found VOI is passed as input to the U-Net in the required input layer dimensions of 96x96(x96)px. We train separate U-Nets for each organ and dimensionality. The implementations are compared by performing five-fold cross-validation on 80 CT images split randomly into training and test sets. For the training of the U-Nets, we use an Nvidia GTX 2080 consumer GPU with 11 GB. We evaluate the segmentation results using various common metrics[8] for semantic segmentation. We also consider the computing time required for training and memory efficiency. The architectural design of both U-Nets is as similar as possible to provide a good basis for comparison, i.e. only the U-Net dimensionality is augmented from 2D to 3D. The basic research question is when to prefer 2D U-Nets. If there is a significant advantage vs. 3D U-Nets, the 2D U-Nets win the competition. Furthermore, conceptually, we lift a Random Regression Forest (RRF) approach from our previous work[9] for organ-specific Bounding Box (BB) detection.

2. MATERIALS AND METHODS

2.1 U-Net Architecture Design

Figure 1 shows the basic architecture of the implemented (using TensorFlow 2.1.0[10]) 2D and 3D U-Net models. The 2D or 3D (denoted in brackets) architectures consist of four down- and up-scaling steps which are concatenated by skip connections (blue arrows). Each yellow box represents a $3x3(x3)$ convolution layer. The number of channels is indicated below the boxes. A $2x2(x2)$ max pooling layer (red) follows each downscale step. Transposed $2x2(x2)$ convolution layers are used (blue) in the up-sampling path. The dropout rates are denoted in front of each dropout layer (brown). A final $1x1(x1)$ convolution with a sigmoidal activation function (magenta) yields the pixel/voxel probabilities for the segmentation. We use zero padding for all convolution layers, except for the final one. The input and output dimensions of the U-Nets are $96x96(x96)$px.

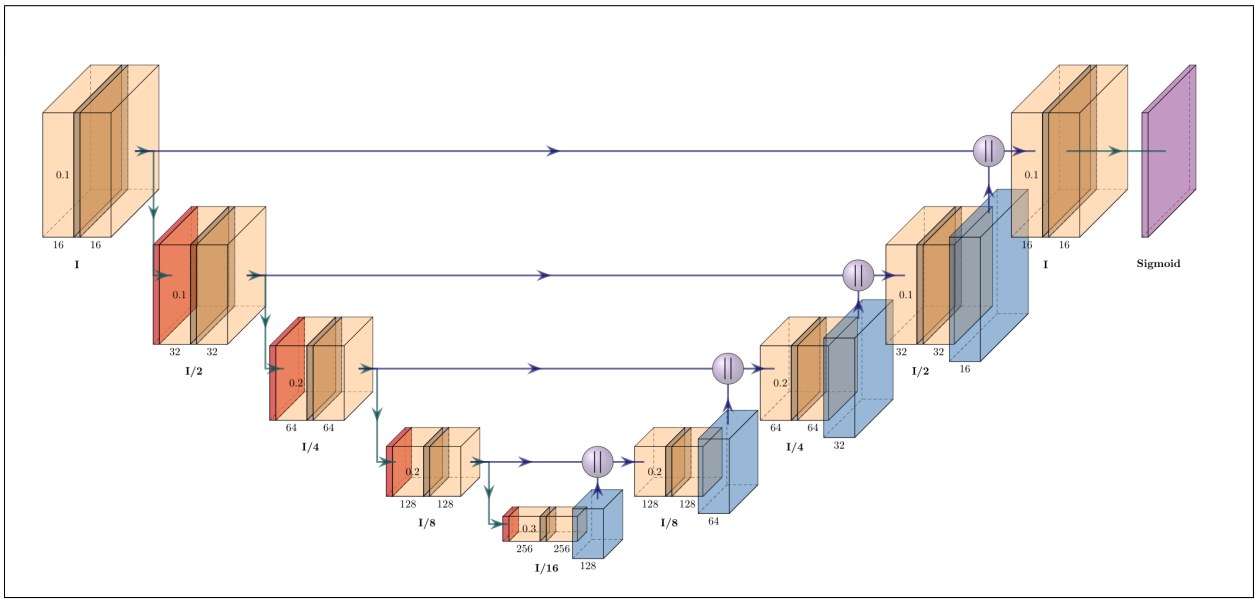


Figure 1. Implemented U-Net Architecture with four down- and up-scaling steps followed by a final sigmoid activation function. Input dimensions of 2D and 3D U-Net are $96x96(x96)$px. Visualization done with PlotNeuralNet.[11]

2.2 Patient Data

The images were obtained from various public sources (VISCERAL, SLIVER07, LiTS, Learn2Reg2020). They consist of 80 abdominal CT scans and corresponding segmentation masks for five target organs (liver, right kidney, left kidney, spleen, pancreas). Some label images were incomplete and therefore had to be completed by two experts (pancreas). The CT scans differ in quality, image noise, field-of-view (fov), contrast agent administration, and slice thickness ($1-5mm$). The covered anatomical area ranges from the axle to the pelvis in most cases. In some CT scans, the visible area is already limited to the abdominal region. The imagery contains patient data with healthy tissue as well as some with pathologies (lesions). As a compromise, the data were resampled to isotropic $2mm^3$ voxels.

2.3 Training and Application

Figure 2 shows the training procedure and the application of a U-Net. In a preceding step, the organs VOI is extracted and its intensity content is transformed into the required input dimensions for the U-Nets. In the second step, the extracted VOI is passed on to either a 2D or a 3D U-Net for semantic segmentation. Since the 2D U-Net naturally cannot handle volumetric data, the problem is treated as a sequential 2D problem. The VOI is split into a sequence of axial slices, which are passed to the 2D U-Net to fit as input. The outputs of the 2D U-net are also 2D images to be re-assembled in 3D. For each organ, one 2D U-Net and one 3D U-Net were trained. The training was conducted over 100 epochs with a batch size of 8 on an Nvidia GTX 2080 with 11 GB. The learning rate was set to the default value of 0.001 using the Adam optimizer. No validation split and no early-stopping conditions were used for training.

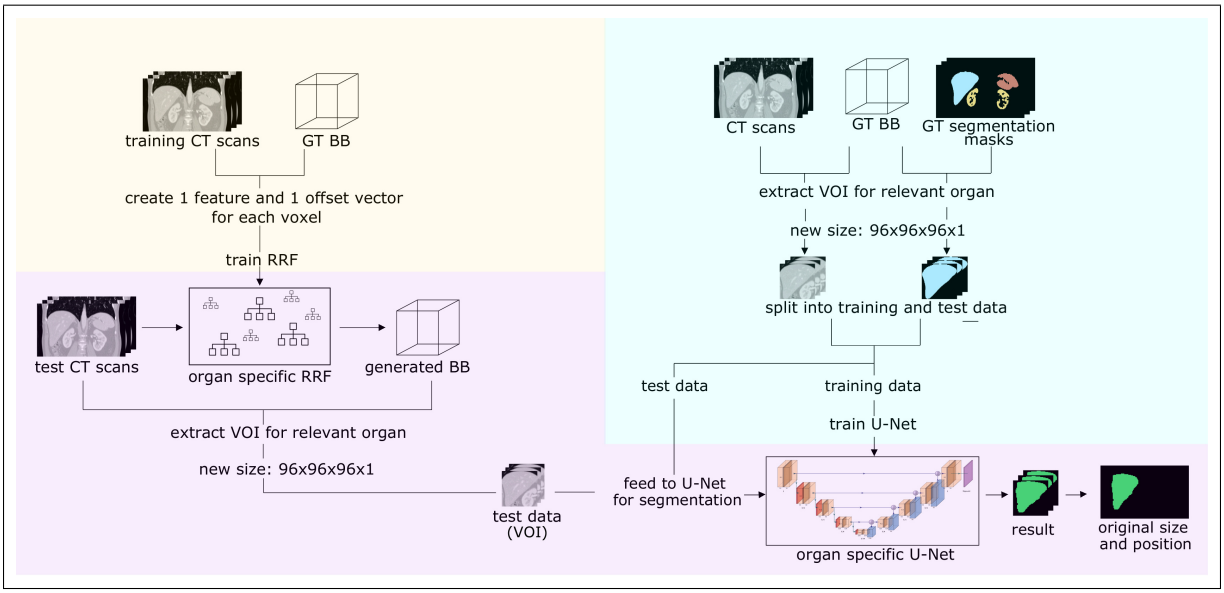


Figure 2. Scheme of methodology for extracting the VOI with RRF generated BB (left in yellow) or using GT BB (right in blue) and framework usage indications (bottom in red).

2.4 Random Regression Forest Bounding Box Detection

In case no Ground Truth (GT) BB data is available for extracting the VOI from the testing data an RRF can be used to generate BB. The left part of Figure 2, highlighted in yellow, shows how the VOI is determined by using a (RRF) detected BB. An approach[9] similar to Criminisi[12] is presented here. Unlike Criminisi, our works[9] train multiple RRF's. To be more precise, one RRF is created per organ. The training of the forests is done as usual. The CT data and the positions of the organ bounding boxes are used as input. The RRF predicts the offset between a voxel and the organ BB. A feature vector and an offset vector are created for the training. We use 50 feature boxes for the creation of the input feature vector. The maximum tree depth is set to 10 to prevent overfitting. Each forest contains exactly 50 trees.

2.5 Evaluation

The procedure using the GT BB for VOI extraction in the first step is the basis for the evaluation. The right part of Figure 2, highlighted in blue, shows how the VOI is extracted without the RFF, only using the GT BB. The size of the extracted volume is then adjusted to the necessary input dimensions to serve as input to the U-Net. The output of the U-net is a pixel/voxel-wise probability map. The values indicate how likely it is that the individual pixel/voxel is part of the organ. To obtain the final organ segmentation, an empirically chosen probability threshold of 0.5 was applied. Only for the low contrasted pancreas, this threshold was set to 0.3, as this yielded better results for both models. Before evaluation takes place, the segmentation result is restored to its original size and placed back at the corresponding position of the previously cut-out VOI. The resulting organ segmentations are eventually compared by their Dice Similarity Coefficients (DSC), Hausdorff Distance (HD), and Average Hausdorff Distance (AVD).[8,13]

HD measures the largest of the directed Hausdorff Distances (see equation 1). The calculation of the directed HD is shown in equation 2 with A and B denoting the surfaces of the two volumes. For each voxel of A, we find its closest voxel from B in the Euclidean metric. The most mismatched voxel of A determines the value of $d(A, B)$.[13] HD targets distance outlier detection, which often occurs in organ segmentation.

$$HD(A, B) = max(d(A, B), d(B, A)) \tag{1}$$

$$d(A, B) = \max_{a \in A} \min_{b \in B} \parallel a - b \parallel \tag{2}$$

We also use the so-called AVD. It is less sensitive to outliers since it takes all distances into account and not only the spotted maximums vs. HD (see equations 3 and 4[8])

$$AVD(A, B) = max(d(A, B), d(B, A)) \tag{3}$$

$$d(A, B) = \frac{1}{N} \sum_{a \in A} \min_{b \in B} \parallel a - b \parallel \tag{4}$$

DSC measures the overlap between two volumes A and B divided by the total number of voxels in both volumes. It is given by equation 5.[14] The maximum achievable value for DSC is 1 (100 %), indicating perfect overlap.

$$DSC = 2 \frac{\mid A \cap B \mid}{\mid A \mid + \mid B \mid} \tag{5}$$

The evaluation of the U-Net models is based on five-fold cross-validation. For this purpose, the 80 images are randomly split into five sets of equal size. Four parts serve as the training set and the remaining part as the test set. This is repeated 5 times until each part has served as test set once. Using 80 volumetric CT patient images, each of the five parts contains 16 images. For a 2D U-Net, in fact, the number of images increases to 1536 axial slice images per part. 2D result labeled sections were re-merged to a 3D volume. Runtime was measured in seconds for each training session. 3D-Slicer 4.10.2[15] was used for figures. SimpleITK 1.2.4[16] was used for the implementation of the evaluation metrics. Statistics, i.e. Shapiro-Wilk pre-tests for normality distribution and Wilcoxon-Signed-Rank (WSR)-tests (per se a paired test), were performed with GNU R 4.0.3.[17]

3. RESULTS

Tables 1, 2, 3, and 4 show the evaluation results for the axial 2D and 3D models. The Shapiro-Wilk-Test for normality was negative for each metric result and organ inhibiting the use of t-tests (paired). Significances, found by the following WSR-test, are marked by the usual stars (p-value $\leq 0.05 \rightarrow {}^{*}$, $\leq 0.01 \rightarrow {}^{**}$, $\leq 0.001 \rightarrow {}^{***}$).Table 1 shows that the HD of the pancreas segmentation is significantly smaller for the 2D U-Net. Also, in terms of AVD (see Table 2), the 2D U-Net architecture performs significantly better for the kidneys. Looking at the DSC evaluation as the main results in Table 3 and the corresponding boxplots in Figure 3, the 2D U-Net is at least on par with the 3D U-Net or even significantly better. The WSR-test has found highly significantly better results for both kidneys in favor of the 2D U-Net ($p \leq 0.001$). It is also significantly better for the liver, however not so striking. In Tables 2 and 3, the standard deviation tends to be larger for the pancreas vs. the rest of the organs for both 2D and 3D models. This is not the case for Table 1. Another summarizing observation is, that particularly for the left kidney, the 3D U-Net seems to be inferior to the 2D U-Net. All metrics (HD, AVD, DSC) reflect this, among them two with significance. Also, compared to the right kidney, the results are consistently weaker, however not significantly.

As can be seen in Table 4, all models need roughly 10 minutes (600 seconds) for training, whereas the 2D U-Nets are slightly faster than the 3D U-Nets in training. Figure 4 shows the DSC of the five-fold cross-validation after training for different numbers of epochs. The 2D U-Net already achieves DSCs greater than 0.9 for most organs after a short training time of 10 epochs. While the 3D U-Net usually needs more epochs to achieve such high DSC results, the 2D U-Net improves faster. After 100 epochs, the 3D model has almost caught up with the 2D model. While all 2D organ models seem to improve or at least stagnate, it can be observed that the DSC of the 2D U-Net of the pancreas worsens slightly.

Figure 5 shows examples of pancreas segmentation results. The GT segmentation is colored in red and the result of the U-Nets is colored in blue. Figure 5(a) shows a segmentation result of the 2D U-Net with a DSC of 0.84. A 2D U-Net segmentation with a very poor DSC of 0.09 is depicted in Figure 5(c). A spatial discontinuity of the slices can be observed in both Figures. Figure 5(d) shows a very poor 3D model segmentation for the same patient (as seen in Figure 5(c)) . In Figure 5(b), the segmentation of the 3D U-Net is very close to the GT segmentation.

Table 1. Mean HD and standard deviation of 2D and 3D U-Net segmentation results for 100 training epochs.

HD					
U-Net	**Liver**	**R. Kidney**	**L. Kidney**	**Spleen**	**Pancreas**
2D	57.00±41.29mm	26.78±26.23mm	26.92±20.36mm	40.12±30.61mm	43.98±21.22mm^{***}
3D	53.85±42.96mm	25.30±25.35mm	29.35±22.82mm	50.13±43.94mm	61.39±31.43mm

Table 2. Mean AVD and standard deviation of 2D and 3D U-Net segmentation results for 100 training epochs.

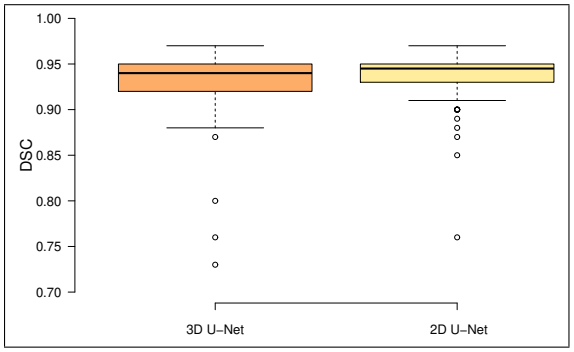
AVD					
U-Net	**Liver**	**R. Kidney**	**L. Kidney**	**Spleen**	**Pancreas**
2D	0.37±0.36mm	0.30±0.30mm^{***}	0.31±0.27mm^{***}	0.33±0.29mm	3.33±2.61mm
3D	0.39±0.43mm	0.36±0.28mm	0.46±0.34mm	0.26±0.17mm	3.15±2.37mm

Table 3. Mean DSC and standard deviation of 2D and 3D U-Net segmentation results for 100 training epochs.

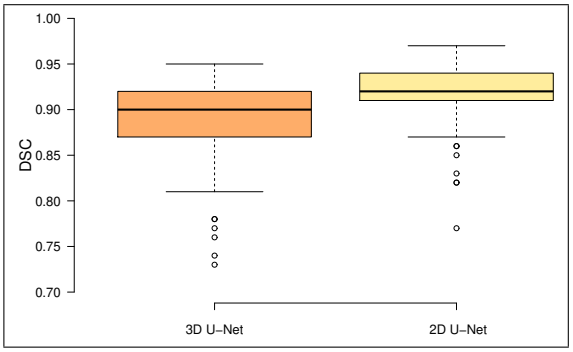
DSC					
U-Net	**Liver**	**R. Kidney**	**L. Kidney**	**Spleen**	**Pancreas**
2D	0.93±0.03*	0.91±0.03***	0.92±0.04***	0.92±0.03	0.56±0.17
3D	0.92±0.03	0.89±0.04	0.85±0.08	0.92±0.03	0.59±0.13

Table 4. Average training time for the 2D and 3D U-Net over 100 training epochs.

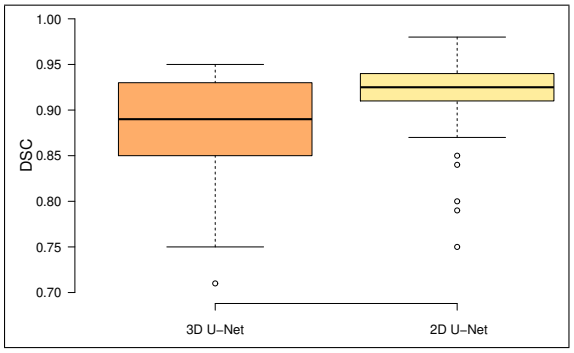
TRAINING TIME					
U-Net	**Liver**	**R. Kidney**	**L. Kidney**	**Spleen**	**Pancreas**
2D	571.72*sec*	572.54*sec*	572.44*sec*	571.84*sec*	573.80*sec*
3D	609.72*sec*	607.56*sec*	607.43*sec*	607.32*sec*	607.81*sec*



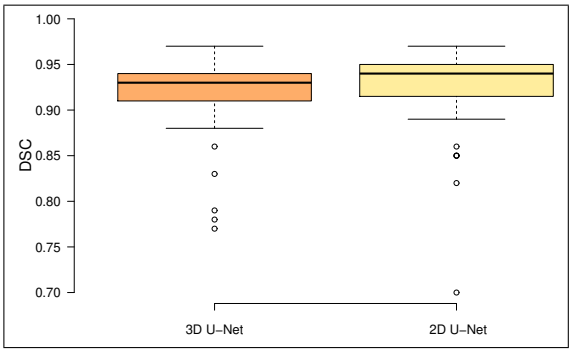
(a) Liver



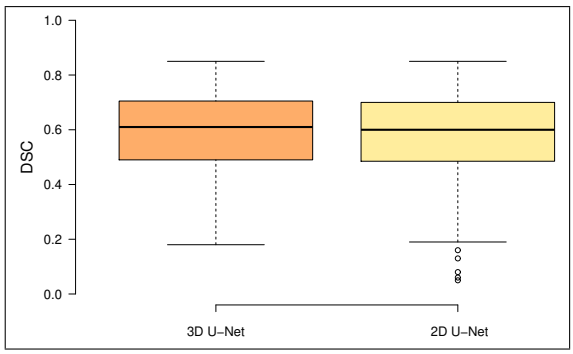
(b) Right kidney



(c) Left kidney

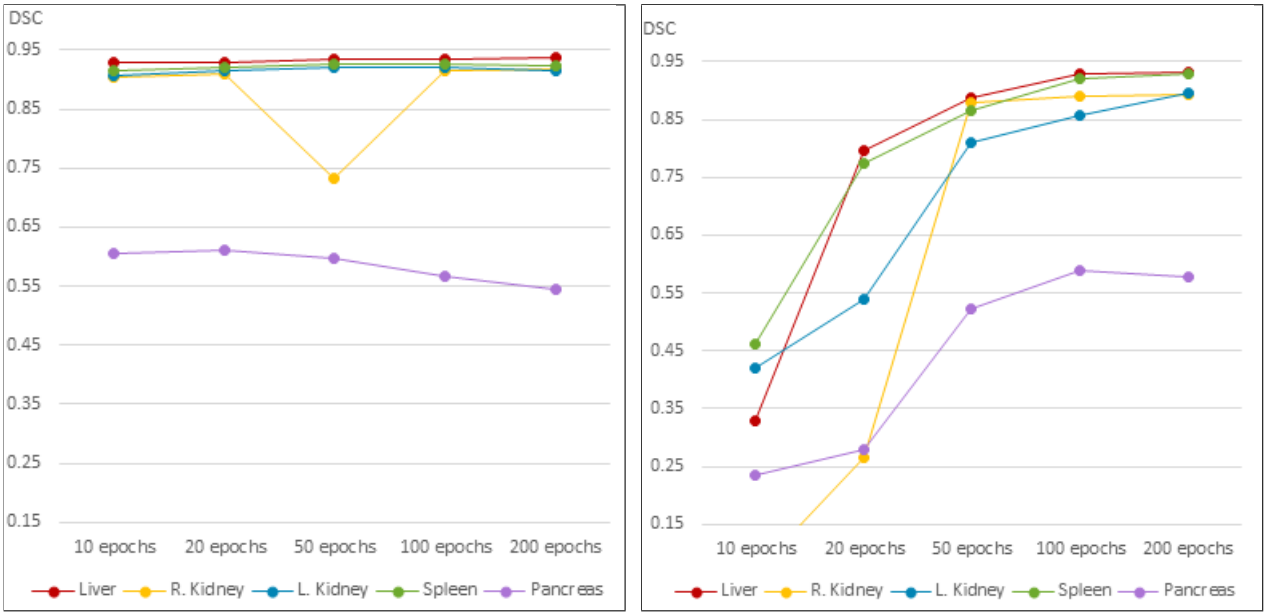


(d) Spleen



(e) Pancreas

Figure 3. Boxplots of all five target organs showing DSC scores for the competing models (100 training epochs). Outliers are marked as dots.



(a) 2D U-Net models

(b) 3D U-Net models

Figure 4. Mean DSC after training 2D and 3D organ models for different numbers of epochs.

4. DISCUSSION AND OUTLOOK

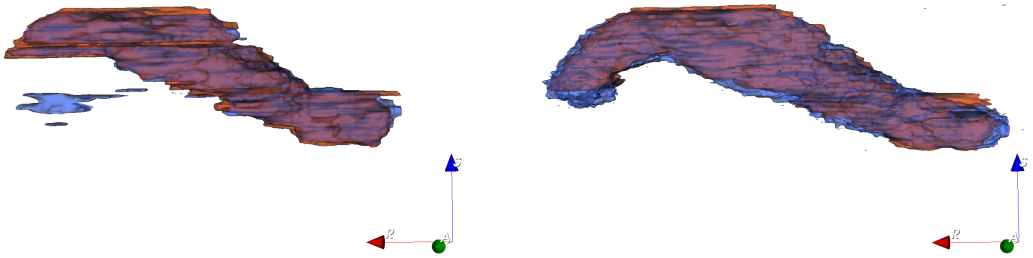
In practice, medical 3D data is often not accessible or available in high quantity. Reasons can be patient privacy protection, radiation dosage, and the costly generation of 3D CT data. The advantage of a 2D network is the larger effective amount of training data that results from splitting the 3D images into stacks of multiple 2D slices. On the other hand, 3D kernels can improve the discriminative power of a network by considering an extra axis. 3D networks, therefore, have a wider field of view with more context available for learning. However, it requires far more computing resources than traditional 2D networks with a 2D kernel.

The measured training time (over 100 epochs) for 2D U-Nets is slightly shorter than for 3D U-Nets, as can be seen in Table 4. Considering that they need fewer epochs to reach relatively high DSC values (Figure 4(a)), their training time until saturation is even shorter. A convergence level is already apparent in the first epochs on the left in Figure 4(a), except for the pancreas. The 2D model for the pancreas peaks after 20 epochs and from then on the DSC gradually worsens. This is an indication that the 2D U-Net is overfitting.

Among all five organs, only the pancreas 3D U-Net delivers slightly better results, in terms of DSC, compared to the 2D model. Additionally, the left boxplot in Figure 3(e) shows no outliers (dots). The pancreas 3D U-Net also achieves segmentation closer to the GT surface, in terms of AVD, and visually (Figure 5(b)). The 2D U-Net suffers from spatial discontinuity of the assembled slices (Figure 5(a) and 5(c)). Interestingly, the 2D U-Net achieves a much smaller HD (Table 1) with a very high significance underlining the robustness of 2D U-Nets in this case. Segmentation of the pancreas in CT is complicated given its shape, oblique orientation, and location. It may be beneficial to first segment other structures surrounding the pancreas. Capturing spatial context with a 3D kernel, as the 3D U-Net does, is also likely to be advantageous in this case. Due to its elongation not dominant on the z-axis, it offers fewer axial slices for training to help the 2D U-Net. Except for HD, it can be concluded that a 3D U-Net offers tendentiously better scores for the pancreas. The interpretation of the HD metric in this study should be under-weighted as it focuses on one outlier for each organ which is highly susceptible to randomness. However, we have observed (Figure 4(a)) that the 2D pancreas U-Net is suffering from overfitting when trained for the full 100 epochs. To make an accurate statement, the metrics of a 2D model trained for fewer epochs should be used for comparison in the future.

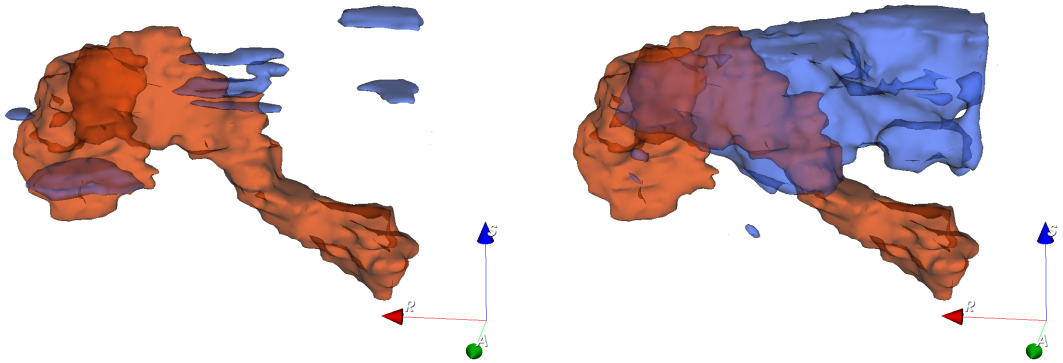
As a take-home message interpreting the main-metric in semantic segmentation evaluation, i.e. the DSC in the first place, the 2D U-Net seems to be much better suited for the kidneys. Clearly, by interpreting the DSC values in Table 3, the 2D U-Net segmentation is highly significantly better. AVD evaluation results indicate the segmentation surface of the kidneys is also highly significantly closer to the GT surface. The kidneys offer a sufficiently high number of axial slices serving as training data for the 2D U-Net. They can be seen as easy structures. The rather round organs are often lighted with contrast agent. Additionally, they are surrounded by fatty tissues with negative CT intensities, protecting them from physical impact. This possibly simplifies segmentation without spatial context (z-dimension) and compensates for the theoretical advantage of the 3D U-Nets with more trainable weights (degrees of freedom). The 2D U-Net is also a significantly better choice for the liver in terms of DSC ($p \leq 0.05$). The DSC of the spleen shows the same values for 2D and 3D U-Net. However, the boxplots for the DSC are to be interpreted slightly in favor of the 2D U-Net. In the case of the spleen, the decision for a better U-Net is therefore not so easy.

In summary and concluding finally, we present a study with an interesting insight into the use of 2D or 3D U-Nets for key-abdominal organs in CT data. A comparison and reasoning about the preferred dimensionality, as well as a decision basis on the selection of a specific U-Net architecture for individual organs, are provided to the reader. The final decision of whether to prefer a 2D or a 3D design is facilitated by giving an overview of which approach performs best for a concrete target organ (limited to the design parameters of this study).



(a) 2D U-Net result with DSC 0.84

(b) 3D U-Net result with DSC 0.84



(c) 2D U-Net result with DSC 0.09

(d) 3D U-Net result with DSC 0.18

Figure 5. Outstanding and poor pancreas segmentation results (blue) with GT segmentation (red). In contrast to the 3D U-Net (right) the 2D U-Net (left) suffers from spatial discontinuity.

Funding: German Research Foundation: DFG MA 6791/1-1, EXPLOR program by Foundation Kessler+Co. for Education and Research: EXPLOR-19AM.

REFERENCES

[1] Mastmeyer, A., Wilms, M., and Handels, H., "Population-based respiratory 4D motion atlas construction and its application for VR simulations of liver punctures," in [*Medical Imaging 2018: Image Processing*], Angelini, E. D. and Landman, B. A., eds., **10574**, 300 – 306, International Society for Optics and Photonics, SPIE (2018).

[2] Mastmeyer, A., Fortmeier, D., and Handels, H., "Random forest classification of large volume structures for visuo-haptic rendering in CT images," in [*Medical Imaging 2016: Image Processing*], Styner, M. A. and Angelini, E. D., eds., **9784**, 670 – 677, International Society for Optics and Photonics, SPIE (2016).

[3] Mastmeyer, A., Wilms, M., Fortmeier, D., Schroder, J., and Handels, H., "Real-time ultrasound simulation for training of us-guided needle insertion in breathing virtual patients," in [*Medicine Meets Virtual Reality 22: NextMed/MMVR22*], **220**, 219, IOS Press (2016).

[4] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*], Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., eds., 234–241, Springer International Publishing, Cham (2015).

[5] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O., "3d u-net: Learning dense volumetric segmentation from sparse annotation," in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*], Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G., and Wells, W., eds., 424–432, Springer International Publishing, Cham (2016).

[6] Meine, H., Chlebus, G., Ghafoorian, M., Endo, I., and Schenk, A., "Comparison of u-net-based convolutional neural networks for liver segmentation in CT," *CoRR* **abs/1810.04017** (2018).

[7] Nemoto, T., Futakami, N., Yagi, M., Kumabe, A., Takeda, A., Kunieda, E., and Shigematsu, N., "Efficacy evaluation of 2D, 3D U-Net semantic segmentation and atlas-based segmentation of normal lungs excluding the trachea and main bronchi," *Journal of Radiation Research* **61**, 257–264 (02 2020).

[8] Taha, A. A. and Hanbury, A., "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging* **15**, 29 (2015).

[9] Mietzner, O. and Mastmeyer, A., "Automatic multi-object organ detection and segmentation in abdominal ct-data," *medRxiv* (2020).

[10] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X., "TensorFlow: Large-scale machine learning on heterogeneous systems," (2015). Software available from tensorflow.org.

[11] Iqbal, H., "Harisiqbal88/plotneuralnet v1.0.0," (2018).

[12] Criminisi, A. and Shotton, J., [*Decision Forests for Computer Vision and Medical Image Analysis*], Springer Publishing Company, Incorporated (2013).

[13] Rucklidge, W., ed., [*The Hausdorff distance*], 27–42, Springer Berlin Heidelberg, Berlin, Heidelberg (1996).

[14] Tustison, N. J. and Gee, J. C., "Introducing dice, jaccard, and other label overlap measures to ITK," *Insight J.* **2** (2009).

[15] Kikinis, R., Pieper, S., and Vosburgh, K., "3d slicer: A platform for subject-specific image analysis, visualization, and clinical support," (2014). Software available from slicer.org.

[16] Lowekamp, B., Chen, D., Ibanez, L., and Blezek, D., "The design of simpleitk," *Frontiers in Neuroinformatics* **7**, 45 (2013).

[17] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2013).