Case Report

Organ Finder – A New AI-based Organ Segmentation Tool for CT

Lars Edenbrandt[1]*, Olof Enqvist[2], Måns Larsson[3], Johannes Ulen[3]

Abstract

Background: Automated organ segmentation in Computed Tomography (CT) is a vital component in many artificial intelligence-based tools in medical imaging. This study presents a new organ segmentation tool called Organ Finder 2.0. In contrast to most existing methods, Organ Finder was trained and evaluated on a rich multi- origin dataset with both contrast and non-contrast studies from different vendors and patient populations.

Approach: A total of 1,171 CT studies from seven different publicly available CT databases were retrospectively included. Twenty CT studies were used as test set and the remaining 1,151 were used to train a convolutional neural network. Twenty-two different organs were studied. Professional annotators segmented a total of 5,826 organs and segmentation quality was assured manually for each of these organs.

Results: Organ Finder showed high agreement with manual segmentations in the test set. The average Dice index over all organs was 0.93 and the same high performance was found for four different subgroups of the test set based on the presence or absence of intravenous and oral contrast.

Conclusions: An AI-based tool can be used to accurately segment organs in both contrast and non-contrast CT studies. The results indicate that a large training set and high-quality manual segmentations should be used to handle common variations in the appearance of clinical CT studies.

Keywords: Image analysis; Artificial intelligence; Machine learning; Computed tomography

Introduction

The application of Artificial Intelligence (AI) in diagnostic imaging has increased drastically during the last decade. At the centre of this development is a type of models called Convolutional Neural Networks (CNN). We have used CNNs in several projects for automated analysis of Positron Emission Tomography (PET)/Computed Tomography (CT), and CT studies. The cornerstone in all these projects has been CNN-based organ segmentation in CT.

Organ segmentation can be used to estimate muscle volume, a measure that can be used to predict survival in cancer patients [1,2]. In PET/CT, CNN-based organ segmentation can be used as a first step in a tumor detection pipeline [3-5]. A CNN trained to detect tumors takes three 3D images as input, the CT, the PET, and a mask created from the organ segmentation. This organ mask provides anatomical localization for a given PET uptake, making correct classification easier. For example, an uptake well inside the urinary

bladder is not likely to be cancer. This approach has been used successfully with, for example, prostate cancer [3], lung cancer [4], and lymphoma [5]. One basic organ segmentation tool from RECOMIA, Segmentation tool v1.0, has been used in several research projects [6]. This tool showed the feasibility of automatically segmenting up to 100 organs in a few minutes. A limitation of the segmentation tool is that only low dose CT without contrast was used for training and testing. This makes the tool perform poorly for organs, such as the kidneys, that look very different in contrast enhanced CT. Hence, the aim of this study was to develop a new CNN for organ segmentation in CT (Organ Finder) using a richer dataset with both non-contrast and contrast studies from different vendors and patient populations. Results will be compared with the performance of Segmentation tool v1.0 [6].

Methods

Patients

We retrospectively included 1,171 CT studies from seven publicly available CT databases (Table 1). These databases include CT studies from over 10 different hospitals with at least four different scanner manufacturers (GE, Philips, Siemens and Toshiba) and varying fields of view (abdomen, chest, torso, from the top of the skull to mid-thigh) [7-19]. Intravenous contrast in different phases (arterial to late phase) was used in roughly 80% of the studies and oral contrast in 35 %, while circa 15% of the studies had no contrast enhancement. A test set was selected to include 20 CT studies (10/10 female/male patients) of which 5 were non-contrast studies, 5 intravenous contrast only, 5 oral contrast only, and 5 with both intravenous and oral contrast. The remaining 1,151 CT studies were used to train the CNN.

Manual segmentations

Professional annotators manually segmented organs using a cloud-based annotation platform. The platform includes basic display features for segmentation of CT studies, for example drawing tools for predefined Hounsfield unit intervals. Every segmentation was first checked by a dedicated quality person and returned to the annotator if adjustments were necessary. A final quality check was done by an experienced nuclear medicine specialist and segmentations with insufficient quality were returned to the annotation team.

Twenty-two different types of organs were segmented, but not all organs were segmented in all CT studies. The left and right kidney, lung, femur, humerus, and hip bone were segmented separately. Left and right clavicles, iliac arteries, and scapulae were combined to one segmentated organ. The ribs and spine were segmented as one organ each. The final database of segmentations consisted of 5,826 quality checked organs (3,856 in intravenous contrast CT and 1,970 in CT without intravenous contrast) (Table 2).

AI tool

Network

The CNN used for organ segmentation is 3D U-Net [20]. It uses convolutions, max pooling and transposed convolutions to process the input CT image on four different resolutions. The information, or features, from lower resolutions are upsampled using transposed convolutions and concatenated with higher resolution features via skip connections across the encoder and decoder part of the CNN, see Figure 1. For an input volume of $100 \times 100 \times 100$ voxels the CNN outputs an estimated class probability for each voxel of a $12 \times 12 \times 12$ volume at the center of the input volume. Before input to the CNN, the CT is resampled to a voxel size of $1 \times 1 \times 3$ mm. The intensities are clamped to $[-800, 800]$ and then rescaled to $[-1, 1]$.

Table 1: Patient material.

Database	Number of Images	Reference
Task 07 Pancreas	420	7
Task 03 Liver	52	7
ACRIN 6668	203	8–10
CT Lymph Nodes	176	8, 11–13
C4KC-KiTS	170	8, 14,15
CT-ORG	117	8, 16–18
Anti-PD-1 Immunotherapy Lung	33	8, 19
Total	**1171**	

Table 2: Segmented organs.

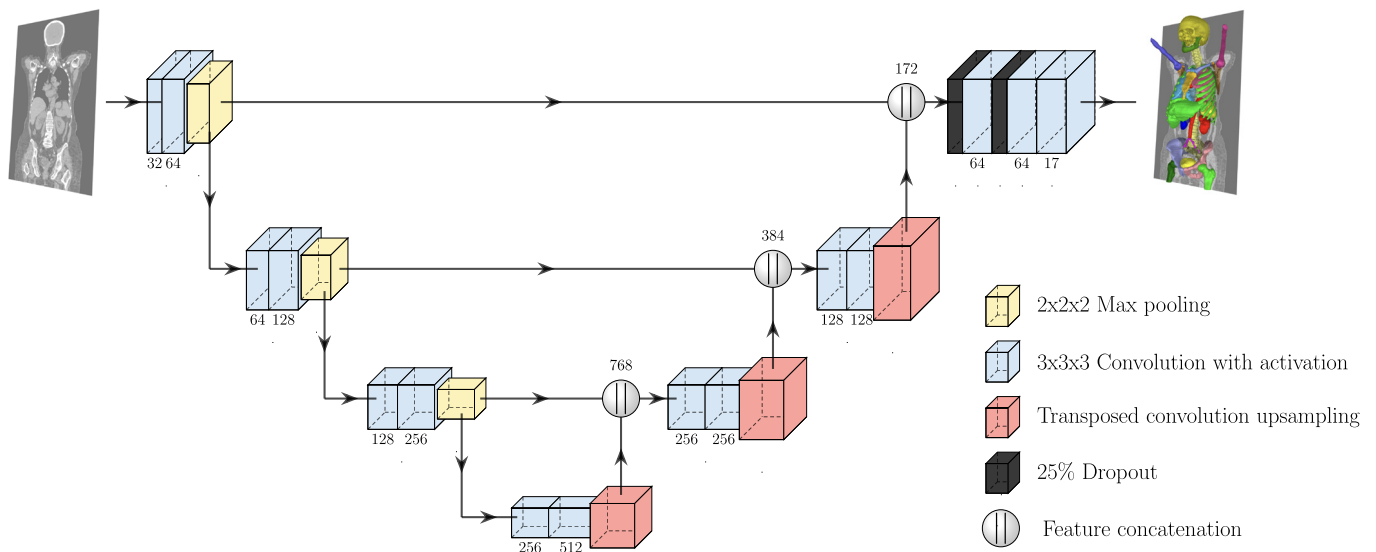| Organ	n
Kidney (left/right)	1158
Liver	604
Spleen	550
Aorta	520
Lung (left/right)	518
Hip bone (left/right)	436
Femur (left/right)	339
Urinary bladder	223
Humerus (left/right)	222
Ribs (combined)	216
Spine	194
Iliac arteries (combined)	169
Clavicles (combined)	157
Sternum	152
Scapulae (combined)	129

Figure 1: Schematic of the 3D U-Net, each layer or operation is represented by a box and connections are visualized with arrows. When applicable, the number of features maps output are shown below (or above) each layer (or operation). All convolutions are done without padding, hence cropping is needed before feature concatenation. Both downsampling with max pooling and upsampling with transposed convolution scales the spatial dimension with a factor two.
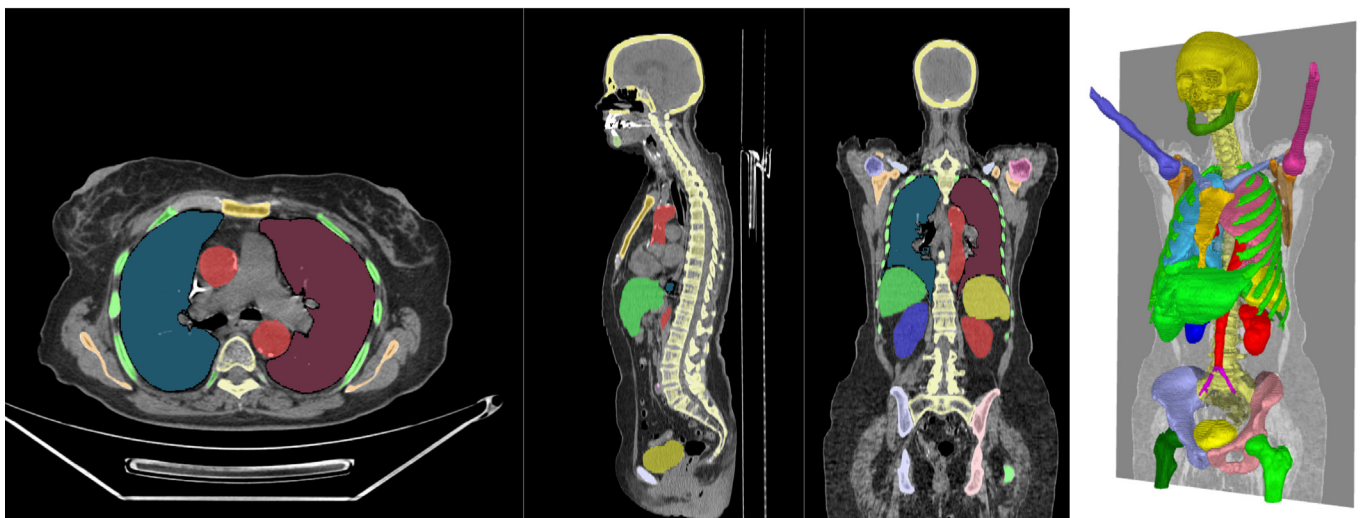

Figure 2: AI-based segmentation of organs in one of the test patients. The average Dice index for all organs was 0.95.

Sampling and training

Training was done with categorical cross-entropy loss and Adam [21] optimization. The learning rate followed the exponential decay schedule with initial learning rate 10^{-4} and a decay rate of 0.95 every 10^5 samples. Sampling was focused on foreground voxels and backgrounds close to foreground since this generally helps the CNN to learn to segment the edges of organs correctly. The different organs were initially sampled evenly but every $5 \cdot 10^5$ training steps, the current model was run on all images in the training set and voxels with high loss was later sampled more frequently.

In addition to applying the cross-entropy loss on the final output layer, deep supervision was used, meaning that the loss was applied on all depth levels of the network. The loss terms were weighted with 1.0, 0.75, 0.5 and 0.25 (highest to lowest scale). For augmentation, a random scaling and rotation was applied to the input and target pair. In addition, Gaussian noise was added to the input images.

Inference

During inference, the CNN, due to being fully convolutional, can be efficiently applied on the entire CT image outputting estimated class probabilities for every voxels. An initial segmentation is then created by taking the argmax for each voxel. Postprocessing is done independently

for each organ. For most organs all segmented voxels except the largest connected component is removed and holes in the segmentation are filled. For iliac arteries, clavicles, scapulae and sternum the two largest components are kept and for the ribs, all components larger than 1 ml are kept.

Statistical Analysis

The AI-based segmentations were compared to the corresponding manual segmentations of the test set. The Sørensen-Dice (Dice) index was used to evaluate the agreement between automated and manual segmentations by analysis of the number of overlapping voxels. For each organ, a (two-sided) sign test was used to test whether there was a significant difference in Dice between Organ Finder and Segmentation tool v1.0.

Results

The new AI-tool for organ segmentation in CT (Organ Finder 2.0, SliceVault AB, Sweden) was applied to the 20 CT studies of the separate test set. One of the test patients is shown in Figure 2. The CNN-based segmentations were compared to the corresponding manual segmentations. Per

Table 3: Dice index per organ in the test set (n=20). Each group (iv+oral, iv, oral, non-contrast consisted of 5 patients (iv-intravenous).

Organ	iv+oral contrast	iv contrast	oral contrast	non-contrast	Total
Aorta	0.93	0.93	0.91	0.90	0.92
Clavicles	0.89	0.87	0.94	0.93	0.91
Femur	0.98	0.98	0.98	0.97	0.98
Hip bone	0.97	0.97	0.97	0.96	0.97
Humerus	0.95	0.97	0.92	0.87	0.92
Iliac arteries	0.80	0.80			0.80
Kidney	0.96	0.95	0.93	0.95	0.95
Liver	0.96	0.97	0.96	0.96	0.96
Lung	0.98	0.98	0.98	0.98	0.98
Mandible			0.89	0.91	0.90
Ribs	0.89	0.90	0.93	0.90	0.90
Scapulae	0.94	0.94	0.95	0.94	0.94
Skull			0.92	0.92	0.92
Spine	0.91	0.94	0.93	0.90	0.92
Sternum	0.93	0.94	0.94	0.82	0.91
Spleen	0.95	0.98	0.95	0.96	0.96
Urinary bladder	0.84	0.75	0.89	0.82	0.82
Total	0.92	0.92	0.94	0.92	0.93

Table 4: Comparison Organ Finder vs. Segmentation tool v1.0 in the test set (n=20) (iv-intravenous).

Organ	Organ Finder	Segmentation tool	Organ Finder	Segmentation tool
Aorta	0.93	0.87	0.91	0.86
Clavicles	0.88	0.49	0.94	0.91
Femur	0.98	0.96	0.97	0.97
Hip bone	0.97	0.93	0.97	0.93
Humerus	0.96	0.34	0.89	0.89
Kidney	0.95	0.61	0.94	0.9
Liver	0.96	0.92	0.96	0.94
Lung	0.98	0.97	0.98	0.97
Mandible			0.9	0.84
Ribs	0.9	0.81	0.91	0.85
Scapulae	0.94	0.84	0.95	0.89
Skull			0.92	0.9
Sternum	0.93	0.82	0.88	0.86
Spleen	0.96	0.79	0.96	0.95
Urinary bladder	0.8	0.7	0.85	0.84
Total	**0.92**	**0.77**	**0.93**	**0.9**

Citation: Edenbrandt L, Enqvist O, Larsson M, Ulen J. Organ Finder – A New AI-based Organ Segmentation Tool for CT. Journal of Radiology and Clinical Imaging 5 (2022): 65-70.

organ metrics are shown in Table 3. The results for the left and right kidneys, lungs, femurs, humeri, and hip bones are grouped. The average Dice index over all organs was 0.93. The AI-tool showed high performance for all four subgroups of the test set based on presence or absence of intravenous and oral contrast. The average Dice index was in the interval 0.90 to 0.98 for all bones and all but two soft tissue organs. The most difficult organs to segment manually, the iliac arteries, showed the lowest Dice index of 0.80 for the CT with intravenous contrast. This organ was not possible to manually segment in the non- contrast CT. Unfortunately, the skull and mandible were not present in the contrast CT studies.

Table 4 shows a comparison between organ segmentation of Organ Finder and Segmentation tool v1.0[6] Segmentation tool v1.0 shows much higher Dice index for the test CT studies without intravenous contrast compared to those with contrast (0.90 vs. 0.77). Organ Finder showed significantly (p<0.01) higher Dice index than Segmentation tool v1.0 for 14 of the 15 organs. The only exception was the urinary bladder having higher Dice index in 13 of the 20 patients (p=0.18).

Discussion and Conclusions

Organ Finder was trained to perform organ segmentation in contrast as well as non-contrast CT. The test results show that this aim was accomplished with average Dice index over all organs of 0.93. The importance of including both contrast and non-contrast studies in a training group was illustrated by the results of Segmentation tool v1.0. The Dice index in the subgroup with intravenous contrast was much lower than the index in the subgroup without intravenous contrast (0.77 vs. 0.90).

Organ Finder was significantly better than Segmentation tool v1.0 even in the subgroup without intravenous contrast (0.93 vs. 0.90). One reason for this increased performance is probably the size difference between the training sets (Organ Finder 1,151 vs. Segmentation tool v1.0 339). A large training set is needed to sufficiently capture variations in appearance due to, for example, patient gender, age, body size and shape, as well as abnormalities due to different diseases. Technical variation, for example, due to presence or absence of intravenous and oral contrast, and differences in field of view should also be represented in a training set. In addition we included CTs from different vendors as well as patient populations from several hospitals in the United States and Europe in the training set.

The superior performance of Organ Finder is most likely also due to higher quality of the manual segmentations. Written specifications on how different organs should be segmented, professional annotators and routines to check segmentation quality for every organ by two dedicated persons contributed to a training set of high quality.

A limitation of Organ Finder is that only 22 organs are included compared to 100 for Segmentation tool v1.0. We decided to limit the number of bones as we found that the value of an AI tool separating individual ribs or vertebrae is limited in most clinical applications. Organ Finder segments all bones from the top of the skull to mid-thigh using 13 labels, which can be compared to the 77 different bones included in Segmentation tool v1.0. Future work with

Organ Finder will include adding more soft tissue organs for example the gastrointestinal tract, heart, muscle, and prostate.

In conclusion, an AI-based tool can be used to accurately segment organs in both contrast and non-contrast CT studies. The results show that a large training set and high-quality manual segmentations should be used to handle common variations in the appearance of a CT in clinical routine.

References

1. Borrelli P, Kaboteh R, Enqvist O, et al. Artificial intelligence-aided CT segmentation for body composition analysis: a validation study. European Radiology Experimental 5 (2021): 1-6

2. Ying T, Borrelli P, Edenbrandt L, et al. Automated artificial intelligence-based analysis of skeletal muscle volume predicts overall survival after cystectomy for urinary bladder cancer. European Radiology Experimental 5 (2021): 1-8.

3. Tra¨ga˚rdh E, Enqvist O, Ule´n J, et al. Freely available, fully automated ai-based analysis of primary tumour and metastases of prostate cancer in whole-body [18f]-psma-1007 Pet CT. Diagnostics 12 (2022).

4. Borrelli P, Go´ngora JLL, Kaboteh R, et al. Freely available convolutional neural network-based quantification of PET/CT lesions is associated with survival in patients with lung cancer. EJNMMI physics 9 (2022): 1-10

5. Sadik M, Lo´pez-Urdaneta J, Ule´n J, et al. Artificial intelligence increases the agreement among physicians classifying focal skeleton/bone marrow uptake in Hodgkin's lymphoma patients staged with [18f] fdg PET/CT— a retrospective study. Nuclear Medicine and Molecular Imaging (2022): 1-7.

6. Tra¨ga˚rdh E, Borrelli P, Kaboteh R, et al. Recomia—a cloud-based platform for artificial intelligence research in nuclear medicine and radiology. EJNMMI physics 7 (2020): 1-12.

7. Antonelli M, Reinke A, Bakas S, et al. The medical segmentation decathlon (2021).

8. Clark KW, Vendt BA, Smith KE, et al. The cancer imaging archive (tcia): Maintaining and operating a public information repository. J Digital Imaging 26 (2013): 1045-1057.

9. Kinahan P, Muzi M, Bialecki B, et al. Data from the ACRIN 6668 Trial NSCLC-FDG-PET (2019).

10. Machtay M, Duan F, Siegel BA, et al. Prediction of survival by [18f] fluorodeoxyglucose positron emission tomography in patients with locally advanced non–small-cell lung cancer undergoing definitive chemoradiation therapy: results of the ACRIN 6668/rtog 0235 trial. Journal of Clinical Oncology 31 (2013): 3823.

11. Roth H, Lu L, Seff A, et al. A new 2.5 d representation for lymph node detection in CT (2015).

12. Roth HR, Lu L, Seff A, et al. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In International Conference on Medical Image Computing and Computer-Assisted Intervention (2014): 520-527.

13. Seff A, Lu L, Cherry KM, et al. 2D View Aggregation for Lymph Node Detection Using a Shallow Hierarchy of Linear Classifiers. In International Conference on Medical Image Computing and Computer-Assisted Intervention (2014): 544-552.

14. Heller N, Sathianathen N, Kalapara A, et al. C4kc Kits Challenge Kidney Tumor Segmentation Dataset (2019).

15. Heller N, Isensee F, Maier-Hein KH, et al. The state of The Art in Kidney and Kidney Tumor Segmentation in Contrast-Enhanced CT Imaging: Results of the Kits19 Challenge. Medical Image Analysis (2021).

16. Rister B, Shivakumar K, Nobashi T, et al. CT-org: A Dataset of CT Volumes with Multiple Organ Segmentations (2019).

17. Rister B, Yi D, Shivakumar K, et al. CT Organ Segmentation Using Gpu Data Augmentation, Unsupervised Labels and Iou loss (2018).

18. Bilic P, Christ PF, Vorontsov E, et al. The Liver Tumor Segmentation Benchmark (lits) (2019).

19. Patnana M, Patel S, Tsao AS. Data from Anti-Pd-1 Immunotherapy Lung (2019).

20. Cicek O, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (2016): 424-432.

21. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. ArXiv preprint ArXiv: 1412.6980 (2014).