3D Bounding Box Detection in Volumetric Medical Image Data: A Systematic Literature Review

Daria Kern and Andre Mastmeyer
Faculty of Optics & Mechatronic, Aalen University, Aalen, Germany
Email: {daria.kern, andre.mastmeyer}@hs-aalen.de

***Abstract*—We analyzed recently published literature from the last five years to identify methods for 3D bounding box detection in volumetric medical image data. A tabular comparison presents the relevant papers and their findings. Various approaches, falling under four identified main categories are described and illustrated. Object detection by means of a 3D bounding box can often be implemented in both 2D and 3D. The advantages and disadvantages of both implementations are discussed. The overview of methods and implementations helps researchers in selecting the most promising approach for their given circumstances. The results show that current research is focused on Deep Learning methods e.g., Convolutional Neural Networks.**

***Index Terms*—literature review, medical imaging, 3D bounding box, object detection, volumetric image data**

I. Introduction

Processing volumetric image data is computationally expensive. In particular, computer-aided diagnosis often requires the processing of volumetric images. The prediction of a 3D Bounding Box (BB) and the extraction of the enclosed Volume of Interest (VOI) is therefore an important preprocessing step for many tasks. E.g., for the classification of malignant tumors or for organ segmentation. Especially, the semantic segmentation of small organs or structures benefits from a prior localization through a 3D BB. Isolated further processing of the VOI facilitates efficiency as irrelevant image areas are excluded and the focus is only on a small, limited image area. Ensuring that only relevant areas need to be processed consequently reduces computing and memory effort. In intervention training and planning, 4D virtual reality simulations are sometimes performed to ensure adequate preparation for medical procedures [1], [2]. The simulations require realistic 3D organ models that need to be reconstructed from patient data. Such data is generated by imaging procedures like Computerized Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Ultrasound (US), or High-Frequency Ultrasound (HFU), just to name a few. 3D organ models can be automatically generated through semantic segmentation of the volumetric image data. Here, too, the preceding extraction of the target organ is advantageous. Adjacent organs are mostly excluded, making the segmentation of the target easier and more efficient. Fig. 1 illustrates the described procedure using a 3D CT scan of the human torso. Displayed are the coronal, sagittal, and axial view with the detected 3D BB around the target organ, which in this case is the liver. In a preceding step, the VOI containing the liver is extracted. In the following steps, the VOI is further processed to retrieve a 3D segmentation of the liver.

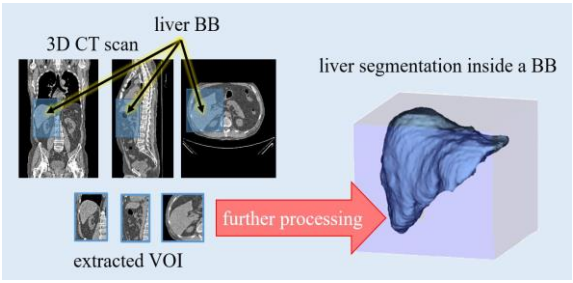

Figure 1. 3D model of the liver inside a bounding box (right). Coronal, sagittal, and axial view of a CT scan with bounding box and extracted volume of interest (left).

In the past, traditional machine learning approaches were often used for predicting a BB around a target. In 2010, Criminisi *et al.* [3] proposed Random Regression Forests (RRFs) to localize target structures in 3D Volumes. Handcrafted features were still required for these approaches. Unlike traditional approaches, modern Deep Learning methods like Convolutional Neural Networks (CNN) do not have to rely on manual feature engineering. CNNs have become immensely popular and have proven to be highly successful in object detection. In 2017, Bob de Vos *et al.* [4] compared CNN based methods to localize anatomical structures in CT scans.

We systematically review papers dealing with 3D BB detection in volumetric medical image data. To capture current trends and developments, the focus is only on recently published papers of the last five years. We distinguish between 2D and 3D implementations and identify four main approaches to predict 3D BBs in volumetric data. The advantages and disadvantages of the implementation options are explained. A tabular presentation of the results provides an overview of our findings.

Manuscript received July 8, 2021; revised December 10, 2021.

doi: 10.18178/joig.10.1.17-27

II. Methodology

In order to find relevant papers, an online search was performed using five well known digital sources. The sources searched were: IEEE Xplore[1], ACM[2], Springer[3], Google Scholar[4] and WoS[5]. The papers of interest deal with methods to predict 3D BBs around targets in volumetric medical image data. The search term was chosen such that papers with point-cloud based data or papers dealing with autonomous driving were excluded. Furthermore, the words "medical" and "localization" had to be included as well as "3D Bounding Box". Thus, the search term was >>"3D Bounding Box" AND "localization" AND medical -vehicle - "point cloud"<< or similar, according to the respective syntax. The search was always limited to publications from 2015 to 2020 to capture recent trends and developments. The search yielded a total of 176 papers, 150 from Google Scholar, 13 from Springer, 7 from IEEE Xplore, and 3 each from WoS and ACM. By abstract screening, a total of 31 papers was selected for the final review. All selected papers are written in English and have been published internationally.

III. Bounding Box Representations

A 3D BB describes a cuboid object in 3D space that ideally fully encloses a target object or structure. In medical image data, this can be an organ or bones, for instance. In contrast to BBs in only two dimensions, an additional dimension, the depth, has to be considered. The area contained in the 3D BB is the VOI. The VOI defines the immediate area in which the target is located. Since we are dealing with 3D data instead of 2D data, it is referred to as volume. 3D BBs can be represented in several ways. Two common ways are the centroid and the two-corner representation, as shown in Fig. 2. In the former, the BB is described by a center point, height, width, and length. The second representation defines two opposite corners. These can be, for example, the minimum and maximum coordinate points.

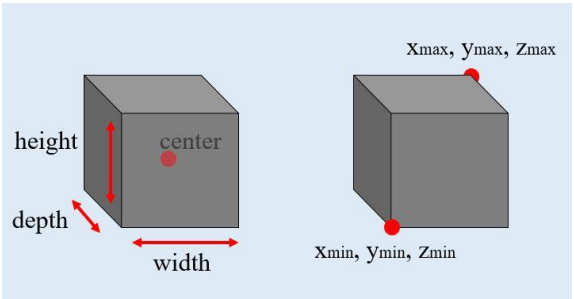

Figure 2. Two different possible BB representations in 3D space. The centroid representation (left) defines a center point and the BB width, height, and depth. The two-corner representation (right) defines the opposite minimum and maximum corners.

[1] ieeexplore.ieee.org
[2] dl.acm.org
[3] link.springer.com
[4] scholar.google.de
[5] webofknowledge.com

IV. 2D vs. 3D Implementation

For object detection in 3D data, a distinction must be made between two different implementation approaches. Meaning, solutions for finding 3D BBs around targets in volumetric medical image data can be implemented in a 3D manner as well as in a 2D manner. A 3D implementation approach simply takes the whole 3D input volume into account. This approach is able to process 3D data. A 2D implementation cannot process 3D data. To still be able to predict a 3D BB around the target, the third dimension first has to be removed. Thus, the 3D volume is split into a stack of multiple, sequential 2D image slices. A distinction is made between different directions in which the 3D volume can be sliced. Slicing usually takes place along one of the three orthogonal image planes (coronal, sagittal, and axial) shown in Fig. 3. in blue, red, and green.

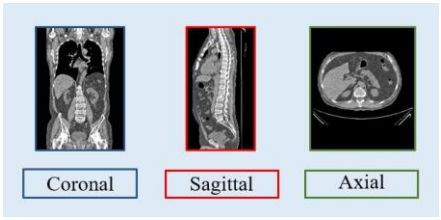

Figure 3. CT scan of the human torso as seen from different viewing directions. The coronal, sagittal, and axial image planes of the 3D volume are indicated as blue, red, and green squares around the CT scan.

Fig. 4 illustrates the 2D image stacks resulting from slicing the volume along the three orthogonal image planes. For better comprehension, the stacks are placed on a 3D model of the human torso. The slicing direction is indicated by the arrows. Along the coronal plane (blue arrow), the torso is sliced from anterior to posterior. Along the sagittal plane (red arrow), from right to left, and along the axial plane (green arrow), from superior to inferior. The resulting stacks of 2D images can be easily processed by 2D implementations. 2D BB detection can take place on the individual 2D image slices. However, after processing is done, any detected 2D BBs in the individual slices would have to be reassembled to form a 3D BB for the volumetric image.

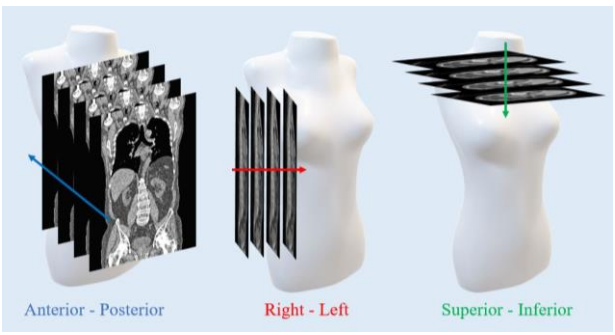

Figure 4. Orthogonal image plane stacks of 2D images resulting from slicing the 3D volume in different directions. Image planes from left to right: coronal, sagittal, axial.

A. Fully 3D Implementation

A 3D implementation approach has the ability to process 3D data and therefore takes the whole image

volume into account. The image volume is composed of many, small volumetric data points, called voxels. Voxels can be seen as the 3D equivalent of pixels as they have one additional dimension along the z-axis, the depth.

Recent work has made extensive use of 3D CNNs [5]-[12] for 3D BB detection. 3D versions of Deep Learning architectures like VGGNet ([13]) [14], Faster R-CNN ([15]) [16], [17] and V-Net ([18]) [19], [20] find great appeal in the research community. A 3D CNN uses a 3D filter kernel instead of a 2D filter kernel. The 3D kernel has to convolve over three axes, the x-axis, the y-axis, and the z-axis. Thus, capturing context information between slices, but also requiring far more resources than its 2D counterpart. Although many approaches today rely on CNNs, more traditional approaches are still present. Y. Zhang *et al.* (2017) [21] extract Haar-like features for every voxel and train a Random Forest to determine a rough 3D BB. R. Gauriau *et al.* (2015) [22] use a cascade of two RRFs.

An advantage of a 3D implementation is, that subsequent reassembly of individual slices is not necessary, reducing the number of work steps required. Although comparisons have shown that 3D approaches generally deliver better results [23]-[25], they still come at a cost. The processing in a 3D manner requires far more computational resources. The advantage of capturing spatial information in all dimensions goes hand in hand with higher memory demand and required computing power. Furthermore, 3D training data is often not available to the same extent as 2D training data. Especially in medical imaging due to patient privacy protection or costly generation, for instance. Thus, 3D approaches often do not have the amount of training data at their disposal that would in fact be needed for optimal training.

B. 2D/2.5D Implementation

A 2D implementation approach does not have the ability to process 3D data as a whole. To solve this problem, the 3D task is transformed into a 2D task. Therefore, the volumetric data is first cut and then examined slice-wise in one of the three orthogonal image planes (i.e., sagittal, coronal, and axial). The 3D image is thus treated as a stack of multiple 2D images and BB detection can be done in 2D. A common approach is to use a single 2D CNN for slice-wise BB detection in either one or all three orthogonal viewing plane directions. A combination of several (usually three) 2D CNNs for all three slicing directions is also possible.

A single 2D CNN to analyze exactly one of the three image plane stacks is implemented in the works of [26]-[30]. 2.5D implementations use adjacent slices as additional channels [31] or dimensions [32] to facilitate the capturing of contextual information, which is usually lost when processing 2D data. The works of [33]-[35], [4] employ a single 2D CNN to analyze all three image plane stacks. Three separate 2D CNNs, one per image plane stack, are implemented in the works of [36]-[40]. G. Humpire-Mamani *et al.* (2018) [41] use a combination of adjacent slices and three separate CNNs. After the data has been processed by either a single 2D model or three 2D models, the results still have to be combined to create a 3D BB. For instance, this can be done by means of a majority voting as implemented by [35]. They first slice the 3D input image in all three viewing plane directions and then a single 2D model processes the input for each direction separately. The output is many different 2D BBs for the target structure in all three directions. The coordinates of the BBs are evaluated together and a majority vote determines the final 3D BB.

A disadvantage of 2D implementation approaches is possibly occurring spatial discontinuity. In 2D BB detection, the 3D image is broken down into individual 2D images and therefore BBs are determined individually for each image. An additional step is required to reassemble the predicted 2D BBs to form a 3D cuboid BB. In doing so, 2D BBs of different shapes and sizes can further complicate the matter. Additionally, spatial discontinuity can occur as seen in Fig. 5. Displayed is a sequence of coronal CT scan slices containing the target organ, the liver. Predicted BBs are displayed as red squares. While the liver was detected in most slices of the sequence, detection failed in one of the middle slices. As the slices are reassembled and a combined 3D BB is to be created, the slice in the middle where no BB was predicted, causes a break in the final 3D BB. Resulting in two separate 3D BB pieces. This problem is irrelevant in a 3D implementation, where the image volume is viewed as a whole and not split. The resulting 3D BB therefore seamlessly encloses the target structure without any breaks. Context information, which a 3D implementation can capture as processing takes place along all three dimensions, is usually also lost in a 2D implementation. However, this problem can be partially solved by considering a couple of adjacent slices as additional channels or dimensions. Big advantages of a 2D approach compared to a 3D approach, are the lower memory consumption and the larger amount of training data. 2D implementation approaches are able to get around the problem of insufficient training data by slicing the few available 3D images into stacks of multiple slices, serving as 2D training data.

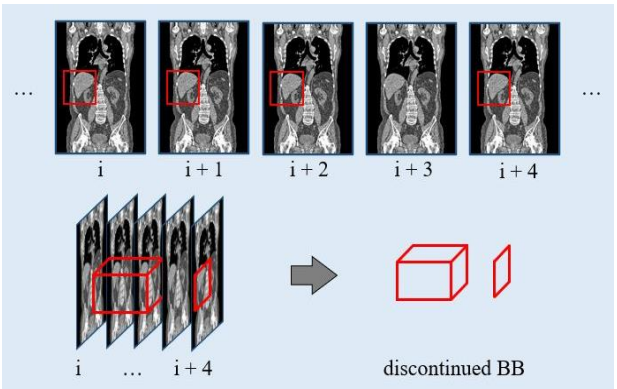

Figure 5. Sequence of coronal slices containing the target (i - i+4) with BBs around the target (red squares in slices i, i + 1, i + 2, and i + 4) and no predicted BB in one of the middle slices (i + 3). Reassembling (bottom left) results in a spatially discontinued 3D BB (bottom right).

V. Identified Approaches

The identified approaches for 3D BB detection approaches in volumetric medical image data are organized into four main categories. The categories are Slice-wise Presence Detection (5 approaches), Coarse Segmentation (9 approaches) and Probability Maps (3 approaches), Anchor-based Approaches (7 approaches), and Deep Reinforcement Learning (3 approaches). Amongst the investigated papers were also approaches that could not be assigned to any of the four main categories. These approaches are listed under the category Other Approaches (4 approaches). Coarse Segmentation and Probability Maps is the most common category with a total of 12 approaches assigned to it. 2D and 2.5D implementations (13 and 3 approaches respectively) are equally as common as 3D implementations (15 approaches). Slice-wise Presence Detection approaches require a pure 2D implementation, whereas approaches of the remaining three main categories can be implemented in both dimensionalities. All approaches of the main categories use CNNs, except two, which fall under the category Coarse Segmentation and Probability Maps and use RRFs or Random Forests instead.

A. Slice-Wise Presence Detection

The approaches of this category are simple but effective. They detect the presence of the target in 2D image slices. Usually in the form of a probability output generated by a CNN. The prediction results of all orthogonal image plane stacks are eventually combined to produce a 3D BB volume. The final 3D BB is therefore the area -or more specifically the volume- where high probability slices from all three stacks overlap. Approaches falling under the slice-wise Presence Detection category work regardless of whether the predictions are done by a single 2D CNN or a combination of three 2D CNNs. However, detecting the presence of an object in each slice requires a 2D implementation. Fig. 6 illustrates a Slice-wise Presence Detection approach with the heart as the target organ. In the first step, the 3D image is cut along each orthogonal image plane to receive three stacks of sequential 2D images. The coronal image plane stack is displayed in blue, the axial stack in green, and the sagittal stack in red. In a second step, each stack is examined slice-wise to determine the slices most probably containing the heart. Only these slices are kept. In the last step, the final 3D BB around the heart is created by considering only the overlapping parts and forming a volume as a result. B. de Vos *et al.* (2016) [36] introduced such an approach using a combination of three CNNs One for each stack. The approach was then adopted by [37], [38]. A combination of three 2.5D CNNs, also considering adjacent slices in addition to the current slice, is successfully used for Slice-wise Presence Detection by [41]. B. de Vos *et al.* (2017) [4] implement a single CNN for all three orthogonal image plane stacks instead and compare the results to those of several other CNN architectures.

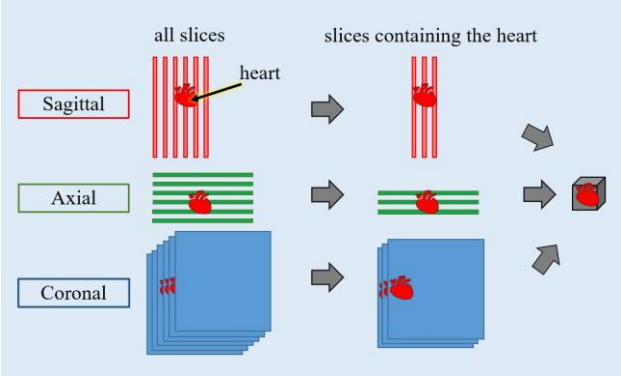

Figure 6. Slice-wise presence detection steps for predicting a BB around the heart. Sequential slices along all three orthogonal image planes are examined and only slices containing the heart are kept. The combined volume yields the final 3D BB.

B. Coarse Segmentation / Probability Maps

The Coarse Segmentation of a target is often an intermediate step for a subsequent, refined segmentation. This means the entire image or volume is processed for segmentation to roughly locate a target. The resulting sub-optimal segmentation is then utilized to place a BB around the area or volume of interest. Fig. 7 illustrates the process step by step, where a Coarse Segmentation (red area) targeting the liver is done on an axial 2D CT scan image. Coarse Segmentation Approaches are implemented by [5], [19], [20], [28], [31]-[35], [40].

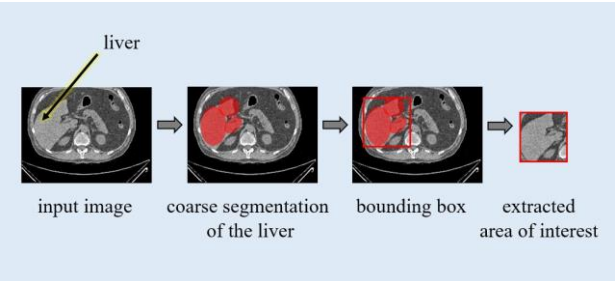

Figure 7. Simplified 2D coarse segmentation procedure, followed by the extraction of the BB area (area in red square) for a possible subsequent segmentation.

Similar to a Coarse Segmentation approach, [39] implement three 2D CNNs for pixel-wise probability detection. The CNNs are applied on the three image plane stacks to obtain confidence heatmaps, which are then used to generate a 3D BB. By applying a threshold against the pixel probabilities, the largest connected component is found and a BB is simply put around it. R. Gauriau *et al.* (2015) [22] calculate voxel probabilities to obtain confidence maps in a 3D manner. They utilize RRFs and divide the localization of several target organs into 2 steps. A first RRF performs a rough localization of all organs at once by creating a global probability map. A second, organ-specific RRF focuses on the individual organs respectively for BB regression. In a similar fashion [21] first take advantage of the knowledge about the relative positions of the target structures and their voxel intensity by using Haar-like features to narrow down the target area. A RRF is then trained on spatial and

intensity features to predict a voxel-wise probability map within the target area. Using a threshold, a BB is placed around the target structure. Coarse Segmentation and Probability Maps implementations are dimension independent.

C. Anchor-Based Approaches

Anchor-based approaches are often used amongst the reviewed works. These approaches have in common that they utilize anchor boxes, which are predefined BB guesses of certain scales and aspect ratios. 2D and 3D anchor boxes are possible. For instance, [29] and [9] follow an Anchor-based approach. Latter combines a 3D CNN and an additional 2D feature extractor for the axial slicing direction to handle various scales and shapes of the target structure. Fig. 8 illustrates three sets of 3D anchor boxes with their anchors (red dots). The anchor boxes are positioned along the x, y, and z-axis of the image volume using their anchors. This is done to propose regions possibly containing the target, which in this case is the liver.

Very popular anchor-based approaches are Faster R-CNN [15] and YOLO [42]. S. Afshari *et al.* (2018) [26] use a modified 2D YOLO to analyze the coronal image plane stack. Whereas YOLO is a one-stage detector, the Faster R-CNN workflow consists of two stages. The backbone network extracts features, which are, together with the anchor boxes, used by a Region Proposal Network (RPN) to generate BB candidates. Instead of positioning the anchors on every single voxel or pixel of the image, anchors are usually positioned in correspondence to voxels or pixels of the encoded feature map. A Fast R-CNN [43] classifier and regressor are then used to determine the class of the object and refine the BBs. K. Chaitanya *et al.* (2020) [17] and [27] use a 3D and 2D Faster R-CNN architecture respectively to detect BBs. X. Xu *et al.* (2019b) [16] modify the 3D Faster R-CNN architecture by removing the classifier and using the Region Proposal Network to propose organ-specific BBs. Relying on the fact that there is at most one instance of an organ, BBs with the same label are merged into one. L. Liu *et al.* (2019) [30] first identify target regions with a Conditional Gaussian Model (CGM) and further localize target structures using a 2D Faster R-CNN.

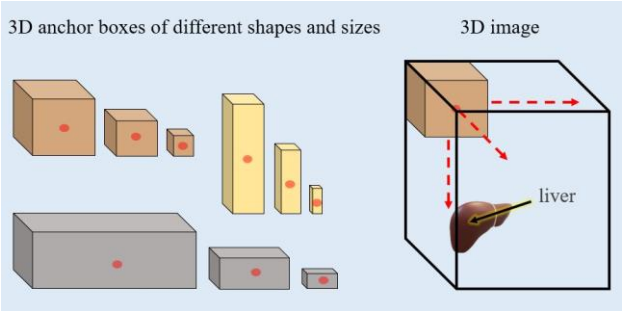

Figure 8. Three 3D sets of anchor boxes (left) propose regions possibly containing the liver in the 3D image volume (right).

D. Deep Reinforcement Learning

Deep Reinforcement Learning (DRL) combines Reinforcement Learning and Deep Learning. In Reinforcement Learning, a so-called agent takes a sequence of actions in order to achieve a certain goal. In doing so, it receives feedback in the form of rewards and penalties. The actions are usually predefined and limited and thereby a final set of elements which the agent chooses from. Through trial and error, the agent tries to maximize the accumulated reward and learns which actions to take. DRL incorporates Deep Neural Networks into this task. The Deep Neural Network analyzes the current state and decides which action to take. The implementation of DRL approaches is dimension independent.

In the work of [10], a 3D CNN receives the current BB voxel values and those of the last four states as input for performing the task of finding the final BB. The actions consider the moving direction, translation, and scaling of the 3D BB. S. Iyer *et al.* (2018, 2020) [7], [12] employ two 3D CNNs for predicting a 3D BB. One CNN for learning the navigation in the coordinate directions and the other to predict the size of the BB dimensions. Starting from a specified point, navigation is allowed to cover 1, 10, or 25 voxels in one direction. DRL is used for this task, replacing random search with guided search using imitation learning. Instead of associating each state with actions that were learned over multiple iterations, imitation learning associates the states with actions that were selected by an expert. Expert decisions are imitated after analyzing the current state. Hence the name imitation learning. Similar to [7], [12], Fig. 9 shows in a simplified way how two 2D Deep Neural Networks (DNNs) can be used to predict a BB. The prediction is performed on a 2D axial CT scan and the target organ is the liver. The first DNN learns how to navigate to the liver and the second DNN learns to determine the size of the BB. Both using a limited set of possible actions, e.g., moving left, right, up or down, increasing or decreasing the BB size.

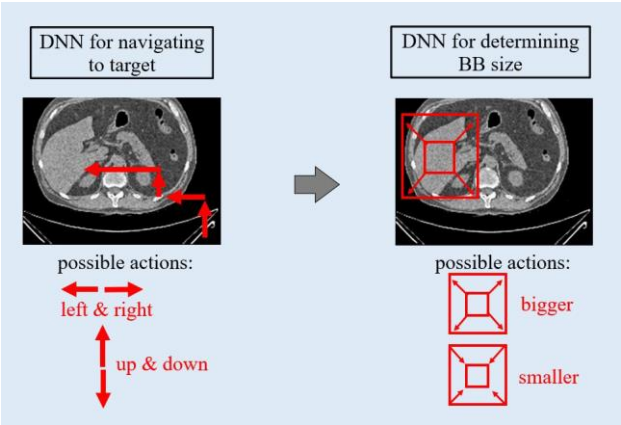

Figure 9. Two Deep Neural Networks (DNNs), in particular CNNs are used for predicting a BB. A first DNN (left) learns how to navigate to the target. A second DNN learns how to determine the size of the BB.

E. Other Approaches

Not all identified approaches did fit into the four main categories. In particular, these were [6], [8], [11], [14]. The approaches are elaborated in the following. S. Han *et al.* (2020) [11] use a 3D modified pre-activation ResNet

[44] for regression on the BB coordinates. Meaning the network outputs 6 coordinates representing start- and endpoints of the 3D BB on each axis. Since the target is framed rather tightly or even partially cut off, they expand the BB symmetrically in each direction to obtain a large enough fixed size BB. R. Janssens *et al.* (2018) [6] also use regression to predict two relative displacement vectors between the two diagonal corners of a BB and a reference voxel. Z. Qiu *et al.* (2018) [14] scan the whole volume using a 3D sliding window, that is large enough to fully contain the target structure. A 10-layer VGGNet [13] serves as the classifier. The sliding window glides through the image and the classifier determines whether it contains the target. In order to train the classifier, examples containing less than 80% of the target are considered as negative and those containing at least 99% as positive. X. Xu *et al.* (2019a) [8] propose a triple-branch fully convolutional network implemented in a 3D manner (see Fig. 10). The input is extended from a single-channel to a three-channel input composed of CT image, enhanced density map, and gradient map. A backbone network produces a feature map which is then used by the branch network to predict the presence of an organ along the axial, coronal, and sagittal direction. The resulting probability curves are then binarized by applying a threshold. The 3D BBs are composed of the largest 1D nonzero component in these three binary curves. At first glance, this approach seems to be similar to the Slice-wise Presence Detection approach. However, in contrast to a 2D implementation, the 3D implementation allows to fully consider spatial context information. Thus, the image is continuously processed as a whole instead of being cut into distinct slices. Through this trick spatial discontinuity could be avoided.

VI. Results

Table I gives an overview of all 31 papers that were considered for the review and finally evaluated. Included are the authors of the paper (Author column) and the imaging procedure by which the volumetric medical image data was produced (Data column). The implementation dimensionality is given (Dim column) as well as the category for 3D BB detection under which the approach falls (Category column). A short description of the procedure gives an insight into the individual work end the employed model architectures (Description column). The target organs or target structures in the body are listed under the Target(s) column. Additionally, a brief summary of the evaluation results of the work is given in the Results column. Metrics used for measurement where mostly Intersection over Union (IoU), Dice Similarity Coefficient (Dice), Average Precision (AP), and Wall Distance (WD). The "Results" column in Table I is non-exhaustive. B. de Vos *et al.* (2017) [4], for instance, did extensive testing and a more detailed evaluation can be found in their paper. Some results are also left blank since no evaluation was performed as localization was a less important intermediate step in the respective works.

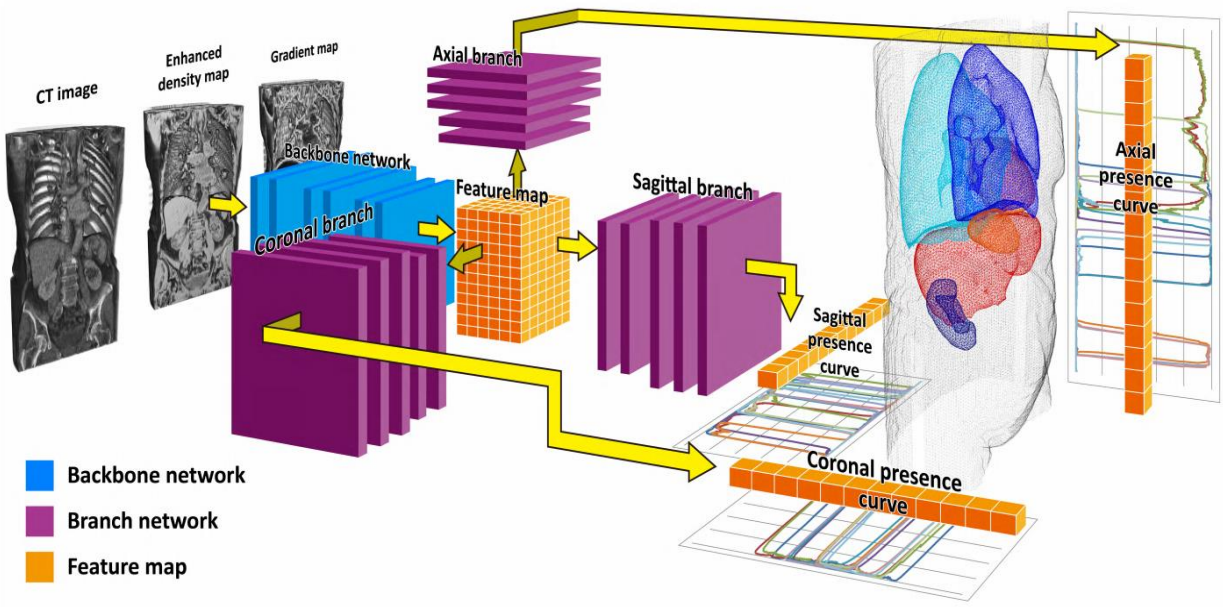

Figure 10. Triple-branch fully convolutional network as proposed by [8]. The 3D input is extended from a single-channel to a three-channel input and fed to the backbone network which extracts the 3D feature map. Three subsequent branch networks synchronously process the extracted 3D feature map and output organ probability curves. On the basis of these curves, the 3D BBs are composed.

TABLE I. LITERATURE FOR 3D BB DETECTION

Author	Data	Dim	Category	Description	Target(s)	Results
R. Gauriau *et al.* (2015) [22]	CT	3D	Probability Map	Two cascaded RRF. 1st RRF for global probability map and 2nd organ specific RRF for local BB improvement	6 abdominal organs	Mean WD 10.7±4 mm, 5.5±4 mm, 5.6±3 mm, 7.9±4 mm, 9.5±4 mm, 13.2±5 mm
B. de Vos *et al.* (2016) [36]	CT	2D	Slice-wise Presence Detection	Combination of three 2D CNNs (AlexNet [45]), each analyzing one orthogonal image plane stack	Heart, aortic arch, d. aorta	Median Dice: 89%, 70%, 85%
M. Zreik *et al.* (2016) [38]	CT	2D	Slice-wise Presence Detection	See Bob D. de Vos *et al.* (2016) [36]	Left ventricle	Complete left ventricle was contained within the BB in all test scans
J. Wolterink *et al.* (2016) [37]	CT	2D	Slice-wise Presence Detection	See Bob D. de Vos *et al.* (2016) [36]	Heart	In all cases the BB contained the whole heart
B. de Vos *et al.* (2017) [4]	CT	2D	Slice-wise Presence Detection	Single 2D CNN (comparing BoBNet [4], VGGNet-16 [13], ResNet-34 [44] and AlexNet [45]) analyzes all three orthogonal image plane stacks	Liver, heart, a. aorta, aortic arch, d. aorta	Dice (comparing CNNs) 96.7%, 96,3%, 96.0%, 95.9%, WD (BoB-Net for liver & heart) 8.87±15.00 mm, 3.11±3.43 mm
Y. Zhang *et al.* (2017) [21]	CT	3D	Probability Map	Combination of 3D Haar-like feature [46] extraction for every voxel and a RF	L. & r. lung, heart	/
H. Roth *et al.* (2018) [39]	CT	2D	Probability Map	Combination of three 2D CNNs (HNN [47]), each analyzing one orthogonal image plane stack	Pancreas	BBs completely surround the pancreases with nearly 100% recall
V. Valindria *et al.* (2018) [5]	MRI	3D	Coarse Segmentation	Weighted 3D CNN, using larger weights for smaller organs	11 abdominal organs, 7 bones	/
M. Tang *et al.* (2018) [29]	US	2D	Anchor Based Approach	Single 2D CNN (VGGNet-16 [13]) analyzes one orthogonal image plane stack	Femoral head	/
R. Huang *et al.* (2018) [40]	US	2D	Coarse Segmentation	Combination of three 2D CNNs (View-based Projection Networks (VP-Nets)), each analyzing one orthogonal image plane stack in real-time	5 key brain structures	Center deviation 1.8±1.4 mm, size difference 1.9±1.5 mm, 3D IoU 63.2±14.7%
S. Afshari *et al.* (2018) [26]	PET	2D	Anchor Based Approach	Single 2D CNN (modified YOLO [42]) analyzes coronal image plane stack	Brain, heart, bladder, r. & l. kidney	Avg. precision 75-98%, recall 94-100%, centroid distance < 14 mm, WD < 24 mm
Z. Qiu *et al.* (2018) [14]	HFU	3D	Other Approach	Sliding window with 3D CNN (10-layer VGGNet [13]) for classification	Brain verticle of embryonic mice	BB containing entire brain verticle 93.7% (single classifier), 96.4% (ensemble of 3 classifiers)
G. Humpire-Mamani *et al.* (2018) [41]	CT	2.5D	Slice-wise Presence Detection	Combination of three 2.5D (adjacent slices) CNNs, each analyzing one orthogonal image plane stack	11 thorax-abdomen organs	Avg. WD of 3.20±7.33 mm, 2nd human observer achieved 1.23±3.39 mm
R. Janssens *et al.* (2018) [6]	CT	3D	Other Approach	3D CNN for regression, predicting two relative displacement vectors between the two diagonal corners of a BB and a reference voxel	Lumbar vertebrae	/
S. Iyer *et al.* (2018) [7]	CT	3D	Deep Reinforcement Learning	Combination of two 3D CNN for DRL and Imitation Learning	Thoracic & lumbar vertebrae	IoU 67.52%, Dice 80.23%
M. Ebner *et al.* (2018) [34] and (2020) [33]	MRI	2D	Coarse Segmentation	Single 2D CNN (P-Net [48]) analyzes all three orthogonal image plane stacks	Fetal brain	IoU 86.54% (normal), 84.74% (presurgical), 83.67 (postsurgical)
X. Wang *et al.* (2019) [28]	US	2D	Coarse Segmentation	Single 2D CNN (U-Net [49]) analyzes one orthogonal image plane stack	Fetal femur	IoU 78.1%
X. Xu *et al.* (2019a) [8]	CT	3D	Other Approach	Single triple-branch 3D CNN with a branch for every orthogonal image plane stack. Additionally, creating a three-channel image as input	11 body organs	IoU 76.44, mean WD 4.36±7.98 mm, mean centroid distance 6.91±9.66 mm
L. Liu *et al.* (2019) [30]	PET /CT	2D	Anchor Based Approach	Combination of a conditional Gaussian model (CGM) and a 2D CNN (Faster R-CNN [15]) for refinement, analyzing one orthogonal image plane stack	Heart, liver, spleen, l. & r. kidney	Center position error thorax: 7.00±2.87 mm (CT), 4.47±2.50 mm (PET) Abdomen: 4.72±2.23 mm (CT), 4.41±2.02 mm (PET)
X. Xu *et al.* (2019b) [16]	CT	3D	Anchor Based Approach	3D CNN (modified Faster R-CNN [15])	11 body organs, 12 head organs	Body: precision 97.91%, recall 98.71%, AP 98.24%, head: 91.11% 91.11%, 84.78%

Y. Wie *et al.* (2019) [9]	CT	3D	Anchor Based Approach	Hybrid multi-atrous and multi-scale network (HMMNet) with multi-atrous3D CNN (MA3DNet) backbone	Liver lesions	Dice 54.8% and 34.2% with IoU of 0.5 and 0.75 respectively
H. Jiang *et al.* (2019) [32]	CT	2.5D	Coarse Segmentation	Single 2.5D (5 adjacent slices, 3D Conv-Kernel) Attention Hybrid Connection Network (AHCNet) analyzes one orthogonal image plane stack.	Liver	/
X. Zhou *et al.* (2019) [35]	CT	2D	Coarse Segmentation	Single 2D CNN analyzes all three orthogonal image plane stacks	17 torso organs	Successfully localized 84.3% (IoU$\geqslant$0.5), mean IoU 70.2%
X. Yang *et al.* (2019) [27]	MRI	2D	Anchor Based Approach	Single 2D CNN (Faster R-CNN [15]) analyzes one orthogonal image plane stack	Left atrium region	100% accuracy
J. Lou *et al.* (2019) [31]	MRI	2.5D	Coarse Segmentation	Single 2.5D (adjacent slices as additional channels) CNN (DS U-Net [50]) analyzes one orthogonal image plane stack	Fetal brain	IoU 91.31±0.08%, centroid distance 2.90±3.53 mm
F. Navarro *et al.* (2020) [10]	CT	3D	Deep Reinforcement Learning	3D CNN (similar to DQN-based network architecture [51]) for DRL	7 abdominal organs	IoU 0.63, abs. median WD 2.25 mm, median distance between centroids 3.65 mm
K. Chaitanya *et al.* (2020) [17]	CT	3D	Anchor Based Approach	3D CNN (Faster R-CNN [15])	Lung nodules	Sensitivity 93% (nodules > 5 mm), 91% (nodules > 3 mm)
S. Han *et al.* (2020) [11]	MRI	3D	Other Approach	3D CNN (modified pre-activation ResNet [52]) for regression on the BB coordinates	Cerebellum	/
T. Xu *et al.* (2020) [19]	HFU	3D	Coarse Segmentation	3D CNN (similar to V-Net [18])	Embryonic mice brain ventricle &body	Dice 81.8%, 91.8%
S. Iyer *et al.* (2020) [12]	CT	3D	Deep Reinforcement Learning	See S. Iyer *et al.* (2018) [7]	Thoracic & lumbar vertebrae	IoU 74/85% (chest), Dice 77/86% (abdomen)
H. Zheng *et al.* (2020) [20]	CT	3D	Coarse Segmentation	Two cascaded 3D CNN (V-Net [18]). 1st CNN for rough localization & 2nd CNN for even more accurate localization	Pancreas	1st & 2nd V-Net Dice 81.38±6.48%, 81.79±7.10%, sensitivity 80.55±9.36%, 81.51±7.22%

VII. Conclusion

We provide a synopsis of the recent works dealing with 3D BB detection in volumetric medical image data. For this purpose, 31 papers of the last 5 years were evaluated. The overview of options presented shall help in selecting a promising approach that also reflects the state of the art in research. The results of the review can also be applied beyond medical imaging as BB detection in 3D data can be applied to other disciplines, too. BB detection helps to save computational cost and to train models for the subsequent semantic segmentation of body areas more specifically, with better results in the end. Some of the presented techniques are also applicable to 2D imagery e.g., detecting, learning and discerning face appearances in photographs [53].

The review differentiates between 3D and 2D implementations, processing the 3D input as a whole or splitting it into several 2D image inputs (slices). The slicing direction is usually along with all or one of the three orthogonal image planes, meaning axial, sagittal or coronal. 2D approaches use either results for a single direction or combine the results of all three directions. It was found 2D implementations are just as common as 3D implementations. 2D approaches have the advantage of generating more training data through slicing. Unfortunately, in contrast to the 3D implementations, they cannot fully capture the spatial context. Among the four identified main categories, Coarse Segmentation and Probability Maps are the most commonly used. Also popular are Anchor-based approaches like Faster R-CNN and YOLO. Furthermore, Deep Reinforcement Learning approaches were identified, as well as Slice-wise Presence Detection of a target structure and finally rather unique approaches e.g., a 3D triple-branch fully convolutional network that simultaneously predicts the presence of an organ in axial, sagittal and coronal direction. It was also found that modern Deep Learning methods, especially with CNNs among them, have largely replaced traditional methods such as RRFs. The most promising and increasingly successful methods seem to be CNNs.

VIII. Future Work

We plan studies in our lab to thoroughly assess the quality and relevance of 3D BB detection for patient modelling in Virtual Reality simulators [54]. We will examine the influence of different imaging modalities [55]-[59] and BB detection quality by Virtual Reality visualization and interaction with detected BBs using haptic force feedback [60]-[63] for quality assurance. In the future, we will also address the accurate and precise BB detection and content segmentation [64] using n-Dimensional image data from various imaging sources. Additionally, the quality of human organ models in the

time-dynamic simulation of 4D medical needle [65], [66] interventions [67], [68] shall profit from the hierarchical and more specific approach.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Daria Kern and Andre Mastmeyer conducted the research, wrote the paper and approved the final version.

ACKNOWLEDGMENT

DFG MA 6791/1-1, EXPLOR-19AM funds granted by Foundation Kessler+Co. for Education and Research.

REFERENCES

[1] D. Fortmeier, M. Wilms, A. Mastmeyer, and H. Handels, "Direct visuo-haptic 4d volume rendering using respiratory motion models," *IEEE Transactions on Haptics*, vol. 8, no. 4, pp. 371-383, 2015.

[2] D. Fortmeier, A. Mastmeyer, J. Schroder, and H. Handels, "A virtual reality system for PTCD simulation using direct visuo-haptic rendering of partially segmented image data," *IEEE J. Biomed. and Health Inform.*, vol. 20, no. 1, pp. 355-366, 2014.

[3] A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu, "Regression forests for efficient anatomy detection and localization in CT studies," in *Proc. International MICCAI Conference on Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging*, 2010, p. 106-117.

[4] B. D. D. Vos, J. M. Wolterink, P. A. D. Jong, T. Leiner, M. A. Viergever, and I. Isgum, "Convnet-based localization of anatomical structures in 3-d medical images," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1470-1481, 2017.

[5] V. V. Valindria, I. Lavdas, J. Cerrolaza, E. O. Aboagye, A. G. Rockall, D. Rueckert, and B. Glocker, "Small organ segmentation in whole-body MRI using a two-stage FCN and weighting schemes," *Lecture Notes in Computer Science*, pp. 346-354, 2018.

[6] R. Janssens, G. Zeng, and G. Zheng, "Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3d fully convolutional networks," in Proc. *IEEE 15th International Symposium on Biomedical Imaging*, Apr. 2018.

[7] S. Iyer, A. Sowmya, A. Blair, C. White, L. Dawes, and D. Moses. (2018). Localization of lumbar and thoracic vertebrae in 3d CT datasets by combining deep reinforcement learning with imitation learning. [Online]. Available: https://cgi.cse.unsw.edu.au/~reports/papers/201803.pdf

[8] X. Xu, F. Zhou, B. Liu, and X. Bai, "Multiple organ localization in CT image using triple-branch fully convolutional networks," *IEEE Access*, vol. 7, pp. 98083-98093, 2019.

[9] Y. Wei, X. Jiang, K. Liu, C. Zhong, Z. Shi, J. Leng, and F. Xu, "A hybrid multi-atrous and multi-scale network for liver lesion detection," in *Proc. International Workshop on Machine Learning in Medical Imaging*, 2019.

[10] F. Navarro, A. Sekuboyina, D. Waldmannstetter, J. C. Peeken, S. E. Combs, and B. H. Menze. Deep reinforcement learning for organ localization in CT. [Online]. Available: https://arxiv.org/abs/2005.04974

[11] S. Han, A. Carass, Y. He, and J. L. Prince, "Automatic cerebellum anatomical parcellation using u-net with locally constrained optimization," *NeuroImage*, vol. 218, p. 116819, 2020.

[12] S. Iyer, A. Sowmya, A. Blair, C. White, L. Dawes, and D. Moses, "A novel approach to vertebral compression fracture detection using imitation learning and patch based convolutional neural network," in *Proc. IEEE 17th International Symposium on Biomedical Imaging*, April 2020, pp. 726-730.

[13] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. [Online]. Available: https://arxiv.org/abs/1409.1556

[14] Z. Qiu, *et al.*, "Deep bv: A fully automated system for brain ventricle localization and segmentation in 3d ultrasound images of embryonic mice," in *Proc. IEEE Signal Processing in Medicine and Biology Symposium*, Dec. 2018.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, Jun. 2017.

[16] X. Xu, F. Zhou, B. Liu, D. Fu, and X. Bai, "Efficient multiple organ localization in CT image using 3d region proposal network," *IEEE Trans. on Med. Imaging*, vol. 38, no. 8, pp. 1885-1898, 2019.

[17] K. C. Kaluva, K. Vaidhya, A. Chunduru, S. Tarai, S. P. P. Nadimpalli, and S. Vaidya, "An automated workflow for lung nodule follow-up recommendation using deep learning," in *Image Analysis and Recognition*, A. Campilho, F. Karray, and Z. Wang, Eds., Cham: Springer Intern. Publ., 2020, pp. 369-377.

[18] F. Milletari, N. Navab, and S. A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. Fourth International Conference on 3D Vision*, Oct. 2016.

[19] T. Xu, *et al.*, "Deep mouse: An end-to-end auto-context refinement framework for brain ventricle and body segmentation in embryonic mice ultrasound volumes," in *Proc. IEEE 17th International Symposium on Biomedical Imaging*, Apr. 2020.

[20] H. Zheng, L. Qian, Y. Qin, Y. Gu, and J. Yang, "Improving the slice interaction of 2.5d CNN for automatic pancreas segmentation," *Medical Physics*, 2020.

[21] Y. Zhang, J. Liu, and J. Liu, "A muti-organ localization method in CT volumes," in *Proc. 9th International Conference on Modelling, Identification and Control*, 2017, pp. 331-335.

[22] R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch, "Multiorgan localization with cascaded global-to-local regression and shape prior," *Medical Image Analysis*, vol. 23, no. 1, pp. 70-83, 2015.

[23] S. Ji, Z. Chi, A. Xu, and Y. Duan, "3d convolutional neural networks for crop classification with multi-temporal remote sensing images," *Remote Sensing*, vol. 10, p. 75, 2018.

[24] X. Zhou, *et al.*, "Performance evaluation of 2D and 3D deep learning approaches for automatic segmentation of multiple organs on CT images," in *Medical Imaging 2018: Computer-Aided Diagnosis*, N. Petrick and K. Mori, Eds., International Society for Optics and Photonics, 2018, vol. 10575, pp. 520-525.

[25] H. Lu, H. Wang, Q. Zhang, S. W. Yoon, and D. Won, "A 3d convolutional neural network for volumetric image semantic segmentation," *Procedia Manufacturing*, vol. 39, pp. 422-428, 2019.

[26] S. Afshari, A. BenTaieb, and G. Hamarneh, "Automatic localization of normal active organs in 3d pet scans," *Comp. Med. Imaging and Graphics*, vol. 70, pp. 111-118, 2018.

[27] X. Yang, N. Wang, Y. Wang, X. Wang, R. Nezafat, D. Ni, and P. Heng. Combating uncertainty with novel losses for automatic left atrium segmentation. [Online]. Available: http://arxiv.org/abs/1812.05807

[28] X. Wang, X. Yang, H. Dou, S. Li, P. A. Heng, and D. Ni, "Joint segmentation and landmark localization of fetal femur in ultrasound volumes," in *Proc. IEEE EMBS International Conference on Biomedical and Health Informatics*, May 2019.

[29] M. Tang, Z. Zhang, D. Cobzas, M. Jagersand, and J. L. Jaremko, "Segmentation-by-Detection: A cascade network for volumetric medical image segmentation," in *Proc. IEEE 15th International Symposium on Biomedical Imaging*, Apr. 2018.

[30] L. Liu, B. Zhang, and H. Wang, "Organ localization in pet/CT images using hierarchical conditional faster r-CNN method," in *Proc. the Third International Symposium on Image Computing and Digital Medicine*, 2019, pp. 249-253.

[31] J. Lou, D. Li, T. D. Bui, F. Zhao, L. Sun, G. Li, and D. Shen, "Automatic fetal brain extraction using multi-stage u-net with deep supervision," in *Machine Learning in Medical Imaging*, H. I. Suk, M. Liu, P. Yan, and C. Lian, Eds., Cham: Springer International Publishing, 2019, pp. 592-600.

[32] H. Jiang, T. Shi, Z. Bai, and L. Huang, "Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in CT volumes," *IEEE Access*, vol. 7, pp. 24898-24909, 2019.

[33] M. Ebner, *et al.*, "An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI," *NeuroImage*, vol. 206, p. 116324, 2020.

[34] M. Ebner, *et al.*, "An automated localization, segmentation and reconstruction framework for fetal brain MRI," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 313-320.
[35] X. Zhou, T. Kojima, S. Wang, X. Zhou, T. Hara, T. Nozaki, M. Matsusako, and H. Fujita, "Automatic anatomy partitioning of the torso region on CT images by using a deep convolutional network with majority voting," in *Medical Imaging 2019: Computer-Aided Diagnosis*, K. Mori and H. K. Hahn, Eds., International Society for Optics and Photonics, 2019, vol. 10950, pp. 256-261.
[36] B. D. D. Vos, J. M. Wolterink, P. A. D. Jong, M. A. Viergever, and I. Isgum, "2D image classification for 3D anatomy localization: employing deep convolutional neural networks," in *Medical Imaging 2016: Image Processing*, M. A. Styner and E. D. Angelini, Eds., International Society for Optics and Photonics, 2016, vol. 9784, pp. 517-523.
[37] J. M. Wolterink, T. Leiner, B. D. D. Vos, R. W. V. Hamersvelt, M. A. Viergever, and I. Isgum, "Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks," *Medical Image Analysis*, vol. 34, pp. 123-136, 2016.
[38] M. Zreik, T. Leiner, B. D. D. Vos, R. W. V. Hamersvelt, M. A. Viergever, and I. Isgum, "Automatic segmentation of ˇ the left ventricle in cardiac CT angiography using convolutional neural networks," in *Proc. IEEE 13th International Symposium on Biomedical Imaging*, 2016, pp. 40-43.
[39] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Medical Image Analysis*, vol. 45, p. 94-107, Apr. 2018.
[40] R. Huang, W. Xie, and J. A. Noble, "VP-Nets: Efficient automatic localization of key brain structures in 3d fetal neurosonography," *Med. Imag. Anal.*, vol. 47, pp. 127-139, 2018.
[41] G. Humpire-Mamani, A. A. A. Setio, B. V. Ginneken, and C. Jacobs, "Efficient organ localization using multi-label convolutional neural networks in thorax-abdomen CT scans," *Physics in Medicine and Biology*, vol. 63, no. 8, p. 085003, 2018.
[42] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Jul. 2017.
[43] R. B. Girshick. Fast R-CNN. [Online]. Available: http://arxiv.org/abs/1504.08083
[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2016.
[45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. the 25th International Conference on Neural Information Processing Systems*, 2012, pp. 1097-1105.
[46] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.
[47] S. Xie and Z. Tu. Holistically-nested edge detection. [Online]. Available: http://arxiv.org/abs/1504.06375
[48] G. Wang, *et al.*, "Deepigeos: A deep interactive geodesic framework for medical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, p. 1559-1572, Jul. 2019.
[49] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234-241.
[50] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P. Heng. 3d deeply supervised network for automatic liver segmentation from CT volumes. [Online]. Available: http://arxiv.org/abs/1607.00582
[51] A. Alansary, *et al.*, "Evaluating reinforcement learning agents for anatomical landmark detection," *Medical Image Analysis*, vol. 53, pp. 156-164, 2019.
[52] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. European Conference on Computer Vision*, 2016, pp. 630-645.
[53] D. Kern, M. Zweng, S. Sello, A. Bagula, and U. Klauck, "Archiving 4.0: Application of image processing and machine learning for the Robben Island Mayibuye Archives," in *Proc. International SAUPEC/RobMech/PRASA Conf.*, 2020, pp. 1-6.
[54] A. Mastmeyer, D. Fortmeier, and H. Handels, "Efficient patient modeling for visuo-haptic VR simulation using a generic patient atlas," *Computer Methods and Programs in Biomedicine*, vol. 132, pp. 161-175, 2016.
[55] P. Zaffino, *et al.*, "Fully automatic catheter segmentation in MRI with 3d convolutional neural networks: Application to MRI-guided gynecologic brachytherapy," *Physics in Medicine & Biology*, vol. 64, no. 16, p. 165008, 2019.
[56] A. Mastmeyer, G. Pernelle, R. Ma, L. Barber, and T. Kapur, "Accurate model-based segmentation of gynecologic brachytherapy catheter collections in MRI-images," *Medical Image Analysis*, vol. 42, pp. 173-188, 2017.
[57] A. Mastmeyer, M. Wilms, D. Fortmeier, J. Schroder, and H. Handels, "Real-Time ultrasound simulation for training of US-guided needle insertion in breathing virtual patients," in *Stud Health Tech and Informatics*, IOS Press, 2016, vol. 220, p. 219.
[58] A. Mastmeyer, D. Fortmeier, and H. Handels, "Random forest classification of large volume structures for visuo-haptic rendering in CT images." in *Proc. SPIE Med. Imag.: Im. Proc.*, 2016, p. 97842H.
[59] A. Mastmeyer, *et al.*, "Model-Based catheter segmentation in MRI-images," in *International Conference on Medi. Imag. Comp. Computer-Assisted Intervention*, 2015.
[60] A. Mastmeyer, D. Fortmeier, and H. Handels, "Evaluation of direct haptic 4d volume rendering of partially segmented data for liver puncture simulation," *Sci. Reps.*, vol. 7, no. 1, pp. 1-15, 2017.
[61] A. Mastmeyer, T. Hecht, D. Fortmeier, and H. Handels, "Ray-Casting based evaluation framework for haptic force feedback during percutaneous transhepatic catheter drainage punctures," *International Journal of Computer Assisted Radiology and Surgery*, vol. 9, no. 3, pp. 421-431, 2014.
[62] D. Fortmeier, A. Mastmeyer, and H. Handels, "GPU-Based visualization of deformable volumetric soft-tissue for real-time simulation of haptic needle insertion," in *German Conf. Med. Imag. Proc.*, Springer, Berlin, Heidelberg, 2012, pp. 117-122.
[63] A. Mastmeyer, D. Fortmeier, and H. Handels, "Direct haptic volume rendering in lumbar puncture simulation," in *Stud Health Technology and Informatics*, IOS Press, 2012, vol. 173, p. 280.
[64] A. Mastmeyer, D. Fortmeier, E. Maghsoudi, M. Simon, and H. Handels, "Patch-Based label fusion using local confidence-measures and weak segmentations." in *Proc. SPIE Medical Imaging: Image Processing*, 2013, p. 86691N.
[65] A. Mastmeyer, T. Hecht, D. Fortmeier, and H. Handels, "Raycasting-Based evaluation framework for needle insertion force feedback algorithms," in *German Conference on Med. Image. Processing*, Springer, Berlin, Heidelberg, 2013, pp. 3-8.
[66] D. Fortmeier, A. Mastmeyer, and H. Handels, "Optimized image-based soft tissue deformation algorithms for visualization of haptic needle insertion." *Studies in Health Technology and Informatics*, vol. 184, p. 136, 2013.
[67] A. Mastmeyer, M. Wilms, and H. Handels, "Population-Based respiratory 4d motion atlas construction and its application for VR simulations of liver punctures," in *Medical Imaging 2018: Image Processing*, International Society for Optics and Photonics, 2018, vol. 10574, p. 1057417.
[68] A. Mastmeyer and M. Wilms, "Interpatient respiratory motion model transfer for virtual reality simulations of liver punctures," *Journal of World Society of Computer Graphics*, vol. 25, no. 1, pp. 1-10, 2017.

Daria Kern is a machine learning research assistant at Aalen University, Germany, where she also received the M.Sc. degree in computer science in 2020. She wrote her master's thesis in cooperation with the University of Western Cape, South Africa. She also has experience in mobile application development as proven by WertheApp. Her current research interests include Deep Learning and Medical Imaging.

Andre Mastmeyer is a Professor at Aalen University, Germany. He holds PhD and Habilitation in Medical Informatics from renowned German universities and gained industry experience in the healthcare sector with IBM Business Consulting and Siemens AG. He was appointed Harvard Medical School Research Fellow in 2014 and has authored 56 peer-reviewed publications in the fields of medical image informatics with artificial intelligence (AI) and n-dimensional, multi-modal data visualization. His scientific and German Research Foundation (DFG) experiences include: After his PhD, he led/leads three successfully managed and conducted or currently running DFG-projects (PI/Co-PI). To this day, he has supervised 22 BSc, MSc and PhD students.