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

Daria Kern [1] and Andre Mastmeyer [2]

[1]Affiliation not available
[2]Aalen University

October 30, 2023

Abstract

This paper discusses current methods and trends for 3D bounding box detection in volumetric medical image data. For this purpose, an overview of relevant papers from recent years is given. 2D and 3D implementations are discussed and compared. Multiple identified approaches for localizing anatomical structures are presented. The results show that most research recently focuses on Deep Learning methods, such as Convolutional Neural Networks vs. methods with manual feature engineering, e.g. Random-Regression-Forests. An overview of bounding box detection options is presented and helps researchers to select the most promising approach for their target objects.

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

Daria Kern (Author)
Faculty of Optics & Mechatronic
Aalen University
Aalen, Germany
daria.kern@hs-aalen.de, hello@dariakern.de

André Mastmeyer (Author)
Faculty of Optics & Mechatronic
Aalen University
Aalen, Germany
andre.mastmeyer@hs-aalen.de

***Abstract*—This paper discusses current methods and trends for 3D bounding box detection in volumetric medical image data. For this purpose, an overview of relevant papers from recent years is given. 2D and 3D implementations are discussed and compared. Multiple identified approaches for localizing anatomical structures are presented. The results show that most research recently focuses on Deep Learning methods, such as Convolutional Neural Networks vs. methods with manual feature engineering, e.g. Random-Regression-Forests. An overview of bounding box detection options is presented and helps researchers to select the most promising approach for their target objects.**

***Keywords*-Literature Review; Medical Imaging; 3D Bounding Box; 3D Object Detection; 3D Object Localization**

I. Introduction

The extraction of a Volume of Interest (VOI) is an important pre-processing step in computer based diagnosis. Tasks such as organ segmentation or classification of malignant tumors usually require a prior localization of the corresponding organ or structure. Especially the semantic segmentation of small organs benefits from a preceding localization. By limiting the data to be examined to a VOI, it is ensured that only relevant areas need to be processed and the computing and memory effort is reduced. For instance in the field of intervention training and planning, 4D Virtual Reality (VR) simulations require realistic 3D patient organ models in order to be an adequate preparation for training and planning medical procedures [1], [2]. The automatic reconstruction of such 3D organ models benefits from the localization of a VOI, as it excludes irrelevant regions and therefore making the segmentation of the relevant structures easier and more efficient.

In this Paper we review 3D Bounding Box (BB) detection in volumetric medical image data. Such data is generated by imaging procedures such as CT (Computerized Tomography), MRI (Magnetic Resonance Imaging), PET (Positron Emission Tomography), US (Ultrasound), HFU (High Frequency Ultrasound), just to name a few. We focus only on recently published papers (last five years) to capture current trends and developments.

II. Methodology

The papers of interest deal with methods to detect 3D BBs around targets in volumetric medical image data. Therefore we used search terms containing *"3D Bounding Box" AND "localization" AND medical -vehicle -"point cloud"* (excluding terms *"vehicle"* and *"point cloud"*) to find relevant papers in public databases and digital libraries. The platforms searched were, IEEE Xplore[1], ACM[2], Springer[3], Google Scholar[4] and WoS[5]. The search was always limited to publications from 2015 to 2020. All papers selected for this review are written in English and have been published internationally. By abstract screening, a total of 31 papers was selected.

III. 3D Bounding Box Representations

A 3D BB describes a cuboid object in 3D space. 3D BBs can be represented in different ways. Two common kinds are the centroid and the two corner representations as seen in Fig. 1. The former defines the center coordinates and the height, width and length of the BB. In the latter case the BB is defined by two opposite corners. Two opposite corners are e.g. the minimum and the maximum coordinate points.

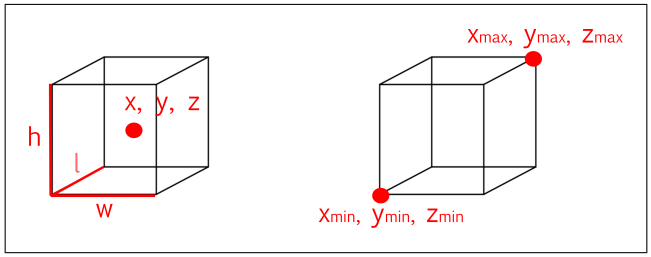

Figure 1. Possible BB representations. Using centroids (left) or two opposite corners (right).

[1] ieeexplore.ieee.org
[2] dl.acm.org
[3] link.springer.com
[4] scholar.google.de
[5] webofknowledge.com, "3D Bounding Box" AND "localization" AND medical NOT vehicle NOT "point cloud", Timespan: Last 5 years.

IV. 2D vs. 3D Implementation

In the past, a popular approach was to train a model using handcrafted features. In 2010 Criminisi [3] proposed Random Regression Forests (RRF) to localize target structures in 3D Volumes. Unlike traditional approaches, modern Deep Leaning methods like Convolutional Neural Networks (CNN) do not have to rely on handcrafted features, but benefit from automated feature extraction. In recent years the focus has clearly shifted towards Deep Learning.

The implementation of solutions for finding 3D BB for target structures in volumetric data can be performed in 2D or 3D. While a 3D implementation takes the whole volume into account, a 2D implementation distinguishes between three orthogonal image planes. These planes are shown in Fig. 2 as red (sagittal), blue (coronal), green (axial) outlined rectangles. In Fig. 2, a 3D BB then is constructed by shifting the colored planes (plane/outside normals pointing away from the patient) around a structure of the human body, e.g. the head.

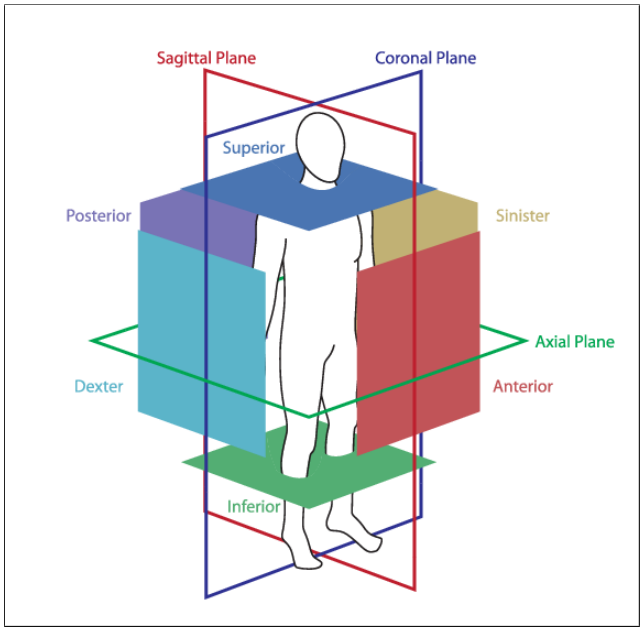

Figure 2. BB walls (6 opaque squares). Sagittal, coronal and axial image viewing planes (outlined rectangles) [4].

A. Fully 3D Implementations

The 3D implementation approach takes the whole 3D image volume as an input to detect a 3D BB. 3D CNNs use 3D instead of 2D filter kernels. The 3D kernel has to convolve over three axes, thus capturing context information between slices, but also requiring far more resources than its 2D counterpart. Recent work has made extensive use of 3D CNNs [5], [6], [7], [8], [9], [10], [11], [12]. 3D versions of Deep Learning architectures like VGGNet([13]) [14], Faster R-CNN([15]) [16], [17] and V-Net ([18]) [19], [20] are very popular. Although most approaches today rely on CNNs, more traditional approaches are still present. Y. Zhang et al. (2017) [21] train a Random Forest after extracting Haar-like features for every voxel to determine a rough 3D BB and R. Gauriau et al. [22] use a cascade of two RRFs.

Although comparisons have shown that 3D approaches generally deliver better results [23], [24], [25], they still come at a cost. The processing in 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.

B. 2D and 2.5D Implementations

The 2D implementation approach deals with 3D localization as a 2D problem. Therefore the volumetric data is 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 several 2D images. A common approach is to use a single 2D CNN or a combination (2.5D) of several (usually three) 2D CNNs for slice wise detection in either one or all three orthogonal viewing plane directions.

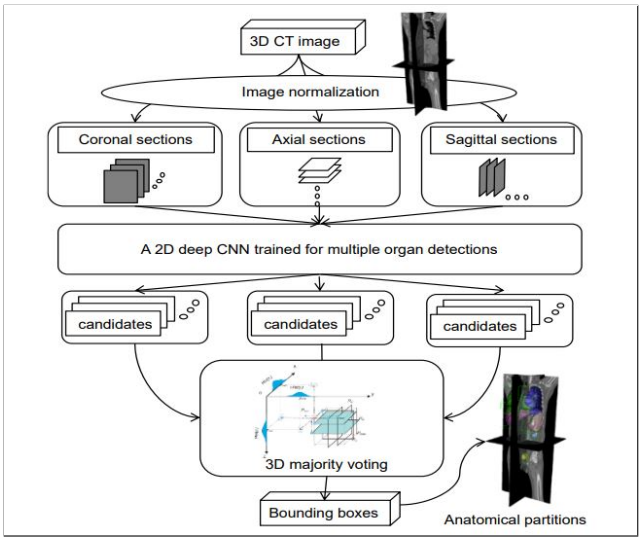

Figure 3. Exemplary process flow for a single 2D CNN combining 3 orthogonal image plane stacks [26].

A single 2D CNN can be implemented to analyze exactly one of the three image plane stacks [27], [28], [29], [30]. Adjacent slices as additional channels [31] or dimensions [32] help to capture contextual information. Another possibility is to analyze all three image planes by using a single 2D CNN three times [33], [34], [26], [4] or three separate 2D CNNs per plane [35], [36], [37], [38], [39]. Adjacent slices and separate CNNs can also be used in combination [40]. After one or more 2D models have processed the data for multiple slicing directions, the results still have to be combined to

create a 3D BB. This can be done by means of a majority voting as seen in Fig. 3. In the illustrated workflow, the 3D input image is sliced in all three viewing plane directions. A single 2D model processes the input for each direction separately. The output are three different BBs for the target structure. The coordinates of the BBs are evaluated together and a majority vote determines the final BB.

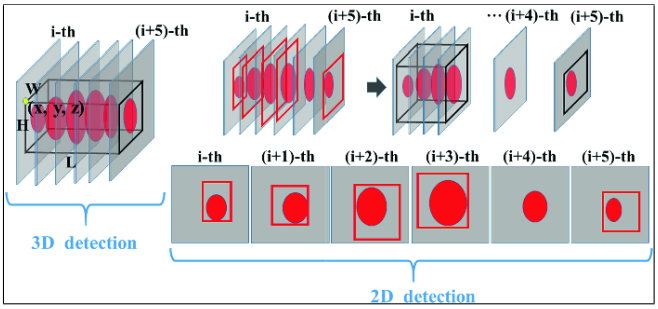

Figure 4. The problem of 2D detection when assembling the slices to form a cuboid BB [9].

The advantage of a 2D compared to a 3D approach, is the lower memory consumption and the larger amount of training data that results from splitting the 3D images into stacks of several slices. A disadvantage is that context information is usually lost. Furthermore, the results of all slices must be assembled to form a cuboid BB, which is further complicated by occurring spatial discontinuity of the slices as seen in Fig. 4. In a 3D detection the image is viewed as a whole. The resulting BB therefore seamlessly encloses the target structure. The problem with 2D detection is that the 3D image is broken down into individual sectional images and BBs are determined individually for each image.

V. Approaches

The following approaches for 3D BB detection in volumetric medical image data have been identified amongst the investigated papers.

A. Slice Wise Box Detection

This approach simply detects the presence of the target structure in every slice. The results for each orthogonal image plane stack are combined to produce a 3D BB [35], [36], [37], [4]. The approach works regardless of whether the results were generated by a single 2D CNN or a combination of three 2D CNN.

B. Coarse Segmentation / Probability Maps

The coarse-segmentation of target structures is often an intermediate step for a subsequent refined segmentation. First, the entire image is viewed to roughly locate one or more targets. The resulting sub-optimal segmentation is then utilized to place a BB around the area of interest [32], [33], [34], [29], [5], [31], [19], [20], [39].

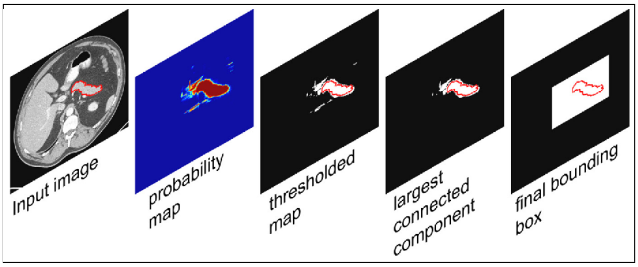

Figure 5. Procedure implemented by H. Roth et al. (2018) [38].

Similar to a coarse-segmentation approach, H. Roth et al. (2018) [38] implement a 2D pixel-wise probability detection in every image plane direction 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. The procedure is shown step by step in Fig.5. R. Gauriau et al. (2015) [22] calculate voxel probabilities to obtain confidence maps in a 3D manner. They utilize RRFs and divide the localization into 2 steps. A first RRF performs a rough localization of all organs at once. A second, organ-specific RRF focuses on the individual organs respectively. In a similar fashion Y. Zhang et al. (2017) [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.

C. Deep Reinforcement Learning

Deep Reinforcement Learning (DRL) combines Reinforcement Learning (RL) and Deep Learning. In RL an agent takes a sequence of actions in order to achieve a certain goal. In doing so, it receives feedback in form of rewards and penalties. Through trial and error, the agent tries to maximize the accumulated reward and learns which actions to take. DRL incorporates Deep Neural Networks (DNN) into this task. The DNN analyzes the current state and decides which action to take. In the work of F. Navarro et al. (2020) [10], the 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, one for learning the navigation in the coordinate directions and the other to predict the size of the BB dimensions.

D. Anchor Based Approaches

Another often seen approach is using anchor boxes, which are predefined BB guesses of certain scales and aspect ratios. For instance M. Tang et al. (2018) [30] and Y. Wei et al. (2019) [9] follow this approach. Latter combine a

3D CNN and an additional 2D feature extractor for the axial slice direction to handle various scales and shapes of the target structure. An output predictor takes the resulting features as input. Very popular anchor-based approaches are Faster R-CNN [15] and YOLO [41]. S. Afshari et al. (2018) [27] 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. A Fast R-CNN [42] classifier and regressor are then used to determine the class of the object and refine the BBs. K. Chaitanya et al. (2020) [17] and X. Yang et al. (2019) [28] use a 3D and 2D Faster R-CNN architecture respectively to detect BBes. 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 a organ, BBs with the same label are merged into one.

Fig. 6 illustrates the common workflow and the one adapted by X. Xu et al. (2019b) [16]. L. Liu et al. (2019) [43] first identify target regions with a Conditional Gaussian Model (CGM) and further localize target structures using a 2D Faster R-CNN.

E. Other Approaches

S. Han et al. (2020) [11] use a 3D modified pre-activation ResNet [44] for regression on the BB coordinates. 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.

X. Xu et al. (2019a) [8] binarize the predicted sagittal, axial and coronal presence probability curves of the target organs by applying a threshold. The 3D BBs are composed by the largest 1D nonzero component in these three binary curves.

VI. Results

Table I gives an overview of the evaluated papers. Included are the author, the image modality, the approach to 3D BB detection, the target structure in the body and the evaluation results of the work. The "Results" column in Table I is nonexhaustive. 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 these papers. Measured was mostly Intersection over Union (IoU), Dice Similarity Coefficient (Dice), Average Precision (AP) and Wall Distance (WD).

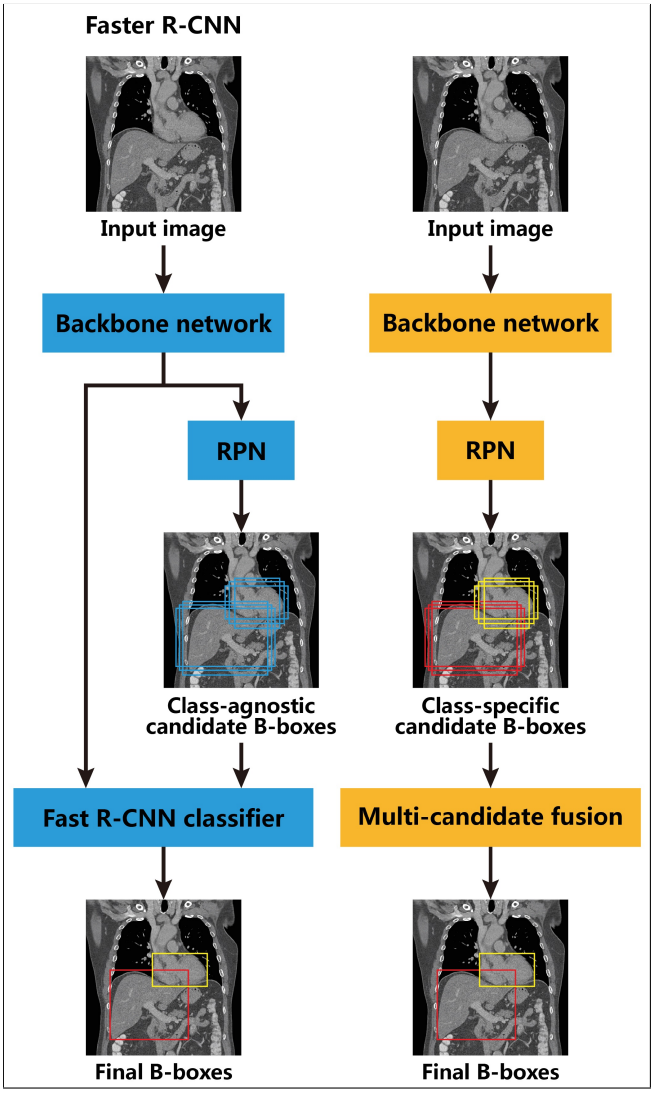

Figure 6. left: general Faster-R-CNN [15] workflow, right: X. Xu et al. (2019b) [16] workflow.

Table I: Literature for 3D BB detection. IoU: Intersection over Union.

Author	Data	Approach	Target(s)	Results
R. Gauriau et al. (2015) [22]	CT	two cascaded RRF. 1st RRF for global coarse segmentation 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) [35]	CT	combination of three 2D CNNs (AlexNet [45]), each analyzing one orthogonal image plane stack	heart, aortic arch, d. aorta	median Dice: 0.89, 0.70, 0.85
M. Zreik et al. (2016) [37]	CT	see Bob D. de Vos et al. (2016) [35]	left ventricle	complete left ventricle was contained within the BB in all test scans
J. Wolterink et al. (2016) [36]	CT	see Bob D. de Vos et al. (2016) [35]	heart	in all cases the BB contained the whole heart
B. de Vos et al. (2017) [4]	CT	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) 0.967, 0.963, 0.960, 0.959, WD (BoBNet for liver and heart) 8.87 ± 15.00 mm, 3.11 ± 3.43 mm
Y. Zhang et al. (2017) [21]	CT	Combination of 3D Haar-like feature [46] extraction for every voxel and a RF	l.&r. lung, heart	/
H. Roth et al. (2018) [38]	CT	combination of three 2D CNNs (HNN [47]), each analyzing one orthogonal image plane stack	Pancreas	BB completely surround the pancreases with nearly 100% recall
V. Valindria et al. (2018) [5]	MRI	weighted 3D CNN for coarse segmentation, using larger weights for smaller organs	11 abdominal organs, 7 bones	/
M. Tang et al. (2018) [30]	US	single 2D CNN (VGGNet-16 [13]) analyzes one orthogonal image plane stack	femoral head	/
R. Huang et al. (2018) [39]	US	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: $63.2 \pm 14.7\%$
S. Afshari et al. (2018) [27]	PET	single 2D CNN (modified YOLO [41]) analyzes coronal image plane stack	brain, heart, bladder, r.&l. kidney	avg. precision 75-98%, recall 94-100%, centroid dist. < 14 mm, WD < 24 mm
Z. Qiu et al. (2018) [14]	HFU	3D CNN (10-layer VGGNet [13])	brain verticle	BB containing entire brain verticle 93.7% (single classifier), 96.4% (ensemble of 3 classifiers)
G. Humpire-Mamani et al. (2018) [40]	CT	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 CNN	lumbar vertebrae	/
S. Iyer et al. (2018) [7]	CT	combination of two 3D CNN for Deep Reinforcement Learning and Imitation Learning	thoracic &lumbar vertebrae	IoU 67.52%, Dice 80.23%
M. Ebner et al.(2018) [34] and (2020) [33]	MRI	single 2D CNN (P-Net [48]) for coarse segmentation analyzes all three orthogonal image plane stacks	fetal brain	IoU 86.54% (normal), 84.74% (presurgical), 83.67% (postsurgical)
X. Wang et al. (2019) [29]	US	single 2D CNN (U-Net [49]) analyzes one orthogonal image plane stack for coarse segmentation	fetal femur	IoU 78.1%

X. Xu et al. (2019a) [8]	CT	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) [43]	PET /CT	combination 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	centre 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 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. Wei et al. (2019) [9]	CT	hybrid multi-atrous and multi-scale network (HMMNet) with multi-atrous 3D 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	single 2.5D (5 adjacent slices, 3D Conv-Kernel) Attention Hybrid Connection Network (AHCNet) for coarse segmentation analyzes one orthogonal image plane stack.	liver	/
X. Zhou et al. (2019) [26]	CT	single 2D CNN analyzes all three orthogonal image plane stacks	17 torso organs	Sucessfully localized 84.3% (IoU ≥ 0.5), mean IoU 70.2%
X. Yang et al. (2019) [28]	MRI	single 2D CNN (Faster-RCNN [15]) analyzes one orthogonal image plane stack	left atrium region	100% accuracy
J. Lou et al. (2019) [31]	MRI	single 2D (adjacent slices as additional channels) CNN (DS U-net [50]) for coarse segmentation analyzes one orthogonal image plane stack	fetal brain	IoU $91.31 \pm 0.08\%$, centroid dist. 2.90 ± 3.53 mm
F. Navarro et al. (2020) [10]	CT	3D CNN (similar to d DQN-based network architecture [51]) for Deep Reinfocement Learning	7 abdominal organs	IoU 0.63, abs. median WD 2.25 mm, median dist. between centroids 3.65 mm
K. Chaitanya et al. (2020) [17]	CT	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 CNN (modified pre-activation ResNet [52])	cerebellum	/
T. Xu et al. (2020) [19]	HFU	3D CNN (similar to V-Net [18]) for coarse segmentation	embryonic mice brain ventricle & body	Dice (coarse segmentation) 0.818, 0.918
S. Iyer et al. (2020) [12]	CT	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	two cascaded 3D CNN (V-Net[18]) for coarse segmentation	pancreas	1st&2nd V-Net Dice: $81.38 \pm 6.48\%$, $81.79 \pm 7.10\%$, sensitivity $80.55 \pm 9.36\%$, $81.51 \pm 7.22\%$

VII. Conclusion and Future Work

We provide a synopsis of the recent works dealing with 3D BB detection in volumetric medical images. For this purpose 31 papers of the last 5 years were evaluated. The review is intended to provide an overview of the current trends as well as information on various options for BB detection in 3D data. 3D and 2D implementations were differentiated, processing the 3D input as a whole or splitting it into several 2D inputs. Various approaches were identified, Coarse Segmentation being the most commonly used. It was also found that Deep Learning methods have largely replaced traditional and other methods, e.g. RRF. The overview of options presented in this review will help future researchers to select a promising approach, which also reflects the state of research. Some of the presented techniques are also applicable to 2D imagery, e.g. detecting, learning and discerning face appearances in photographs [53]. Traditional techniques such as RRFs have been augmented by Deep Learning techniques, especially with CNNs among them. The most promising and increasingly successful methods seem to be CNNs, as they combine traditional signal processing approaches (convolution filtering) with automatic learning from examples in Neural Networks. 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.

To assess the quality and relevance of BB detection for patient modelling in VR simulators [54], [1], [2] thoroughly, we plan studies in our lab to examine the influence of different imaging modalities [55], [56], [57], [58], [59] and BB detection quality by VR visualization and interaction with detected BBs using haptic force feedback [60], [58], [61], [62], [63] for quality assurance. In the future, we will also address the accurate and precise BB detection and content segmentation [64] using nD image data from various imaging sources. Additionally the quality of 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.

Acknowledgment

German Research Foundation 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 - TOH*, vol. 8, no. 4, pp. 371–383, 2015.

[2] D. Fortmeier, A. Mastmeyer, J. Schröder, and H. Handels, "A virtual reality system for ptcd simulation using direct visuo-haptic rendering of partially segmented image data," *IEEE Journal of Biomedical and Health Informatics - JBHI*, 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 *Proceedings of the 2010 International MICCAI Conference on Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging*, ser. MCV'10. Berlin, Heidelberg: Springer-Verlag, 2010, p. 106–117.

[4] B. D. de Vos, J. M. Wolterink, P. A. de Jong, T. Leiner, M. A. Viergever, and I. Išgum, "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*, p. 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," *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Apr 2018. [Online]. Available: http://dx.doi.org/10.1109/ISBI.2018.8363715

[7] S. Iyer, A. Sowmya, A. Blair, C. White, L. Dawes, and D. Moses, "Localization of lumbar and thoracic vertebrae in 3d ct datasets by combining deep reinforcement learning with imitation learning," 2018.

[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. 98 083–98 093, 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 *Machine Learning in Medical Imaging. MLMI 2019*, vol. 11861. Springer, 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," 2020.

[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 *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, April 2020, pp. 726–730.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.

[14] Z. Qiu, J. Langerman, N. Nair, O. Aristizabal, J. Mamou, D. H. Turnbull, J. Ketterling, and Y. Wang, "Deep bv: A fully automated system for brain ventricle localization and segmentation in 3d ultrasound images of embryonic mice," *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, Dec 2018. [Online]. Available: http://dx.doi.org/10.1109/SPMB.2018.8615610

[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, p. 1137–1149, Jun 2017. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2016.2577031

[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 Transactions on Medical 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 International Publishing, 2020, pp. 369–377.

[18] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *2016 Fourth International Conference on 3D Vision (3DV)*, Oct 2016. [Online]. Available: http://dx.doi.org/10.1109/3DV.2016.79

[19] T. Xu, Z. Qiu, W. Das, C. Wang, J. Langerman, N. Nair, O. Aristizabal, J. Mamou, D. H. Turnbull, J. A. Ketterling, and et al., "Deep mouse: An end-to-end auto-context refinement framework for brain ventricle and body segmentation in embryonic mice ultrasound volumes," *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Apr 2020. [Online]. Available: http://dx.doi.org/10.1109/ISBI45749.2020.9098387

[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 *2017 9th International Conference on Modelling, Identification and Control (ICMIC)*, 2017, pp. 331–335.

[22] R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch, "Multi-organ 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, 01 2018.

[24] Z. Xiangrong, Y. Kuzuma, T. Ryosuke, Z. Xinxin, H. Takeshi, F. Hiroshi, W. Song, and K. Takuya, "Performance evaluation of 2d and 3d deep learning approaches for automatic segmentation of multiple organs on ct images," p. 105752C, 2018.

[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, 25th International Conference on Production Research Manufacturing Innovation: Cyber Physical Manufacturing August 9-14, 2019 — Chicago, Illinois (USA).

[26] 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., vol. 10950, International Society for Optics and Photonics. SPIE, 2019, pp. 256 – 261. [Online]. Available: https://doi.org/10.1117/12.2512651

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

[28] X. Yang, N. Wang, Y. Wang, X. Wang, R. Nezafat, D. Ni, and P.-A. Heng, "Combating uncertainty with novel losses for automatic left atrium segmentation," *Lecture Notes in Computer Science*, p. 246–254, 2019.

[29] 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," *2019 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, May 2019. [Online]. Available: http://dx.doi.org/10.1109/BHI.2019.8834615

[30] M. Tang, Z. Zhang, D. Cobzas, M. Jagersand, and J. L. Jaremko, "Segmentation-by-detection: A cascade network for volumetric medical image segmentation," *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Apr 2018. [Online]. Available: http://dx.doi.org/10.1109/ISBI.2018.8363823

[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. 24 898–24 909, 2019.

[33] M. Ebner, G. Wang, W. Li, M. Aertsen, P. A. Patel, R. Aughwane, A. Melbourne, T. Doel, S. Dymarkowski, P. De Coppi, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain mri," *NeuroImage*, vol. 206, p. 116324, 2020.

[34] M. Ebner, G. Wang, W. Li, M. Aertsen, P. Patel, R. Aughwane, A. Melbourne, T. Doel, A. David, J. Deprest, S. Ourselin, and T. Vercauteren, *An Automated Localization, Segmentation and Reconstruction Framework for Fetal Brain MRI*. Springer International Publishing, 09 2018, pp. 313–320.

[35] B. D. de Vos, J. M. Wolterink, P. A. de Jong, M. A. Viergever, and I. Išgum, "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., vol. 9784, International Society for Optics and Photonics. SPIE, 2016, pp. 517 – 523. [Online]. Available: https://doi.org/10.1117/12.2216971

[36] J. M. Wolterink, T. Leiner, B. D. de Vos, R. W. van Hamersvelt, M. A. Viergever, and I. Išgum, "Automatic coronary artery calcium scoring in cardiac ct angiography using paired convolutional neural networks," *Medical Image Analysis*, vol. 34, pp. 123 – 136, 2016, special Issue on the 2015 Conference on Medical Image Computing and Computer Assisted Intervention.

[37] M. Zreik, T. Leiner, B. D. de Vos, R. W. van Hamersvelt, M. A. Viergever, and I. Išgum, "Automatic segmentation of the left ventricle in cardiac ct angiography using convolutional neural networks," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 40–43.

[38] 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. [Online]. Available: http://dx.doi.org/10.1016/j.media.2018.01.006

[39] R. Huang, W. Xie, and J. Alison Noble, "Vp-nets : Efficient automatic localization of key brain structures in 3d fetal neurosonography," *Medical Image Analysis*, vol. 47, pp. 127 – 139, 2018.

[40] G. Humpire-Mamani, A. A. A. Setio, B. van 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 8, p. 085003, 2018.

[41] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2017.690

[42] R. B. Girshick, "Fast R-CNN," *CoRR*, vol. abs/1504.08083, 2015. [Online]. Available: http://arxiv.org/abs/1504.08083

[43] L. Liu, B. Zhang, and H. Wang, "Organ localization in pet/ct images using hierarchical conditional faster r-cnn method," New York, NY, USA, p. 249–253, 2019. [Online]. Available: https://doi.org/10.1145/3364836.3364886

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016. [Online]. Available: http://dx.doi.org/10.1109/cvpr.2016.90

[45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. Red Hook, NY, USA: Curran Associates Inc., 2012, p. 1097–1105.

[46] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–I.

[47] S. Xie and Z. Tu, "Holistically-nested edge detection," *CoRR*, vol. abs/1504.06375, 2015. [Online]. Available: http://arxiv.org/abs/1504.06375

[48] G. Wang, M. A. Zuluaga, W. Li, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and 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. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2018.2840695

[49] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, p. 234–241, 2015.

[50] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3d deeply supervised network for automatic liver segmentation from ct volumes," *Lecture Notes in Computer Science*, p. 149–157, 2016.

[51] A. Alansary, O. Oktay, Y. Li, L. L. Folgoc, B. Hou, G. Vaillant, K. Kamnitsas, A. Vlontzos, B. Glocker, B. Kainz, and D. Rueckert, "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 *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 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 *2020 International SAUPEC/RobMech/PRASA Conference*, 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 - CMPB*, vol. 132, pp. 161–175, 2016.

[55] P. Zaffino, G. Pernelle, A. Mastmeyer, A. Mehrtash, H. Zhang, R. Kikinis, T. Kapur, and M. F. Spadea, "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. Schröder, and H. Handels, "Real-time ultrasound simulation for training of us-guided needle insertion in breathing virtual patients," in *Studies in Health Technology and Informatics*. 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 Medical Imaging: Image Processing*, 2016, p. 97842H.

[59] A. Mastmeyer, G. Pernelle, L. Barber, S. Pieper, D. Fortmeier, S. Wells, H. Handels, and T. Kapur, "Model-based catheter segmentation in mri-images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI*, 2015.

[60] A. Mastmeyer, D. Fortmeier, and H. Handels, "Evaluation of direct haptic 4d volume rendering of partially segmented data for liver puncture simulation," *Scientific Reports*, 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 - IJCARS*, 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 Conference on Medical Image Processing - BVM*. Springer, Berlin, Heidelberg, 2012, pp. 117–122.

[63] A. Mastmeyer, D. Fortmeier, and H. Handels, "Direct haptic volume rendering in lumbar puncture simulation," in *Studies in 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, "Ray-casting-based evaluation framework for needle insertion force feedback algorithms," in *German Conference on Medical Image Processing - BVM*. 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 *Proc. SPIE Medical Imaging: Image Processing*, vol. 10574. International Society for Optics and Photonics, 2018, 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 - WSCG*, vol. 25, no. 1, pp. 1–10, 2017.