One-stage Multi-task Detector for 3D Cardiac MR Imaging

Weizeng Lu[1], Xi Jia[2], Wei Chen[2], Nicolo Savioli [3], Antonio de Marvao[3],
Linlin Shen[1*], Declan O'Regan[3] and Jinming Duan [2]

[1]Computer Vision Institute, School of Computer Science and Software Engineering,
Shenzhen University, Shenzhen, China

[2]School of Computer Science, University of Birmingham, United Kingdom

[3]MRC London Institute of Medical Sciences, Imperial College London, United Kingdom

Abstract

In this paper, we propose a novel multi-task learning framework, for real-time, simultaneous landmark location and bounding box detection in 3D space. Our method extends the famous single shot multi box detector (SSD) from single-task learning to multitask learning and from 2D to 3D. Furthermore, we propose a post-processing approach to refine the network landmark output, by averaging the candidate landmarks. Owing to these settings, the proposed framework is fast and accurate. For 3D cardiac magnetic resonance (MR) images with size 224×224×64, our framework runs ~128 volumes per second (VPS) on GPU and achieves 6.75mm average point-to-point distance error for landmark location, which outperforms both state-of-the-art and baseline methods. We also show that segmenting the 3D image cropped with the bounding box results in both improved performance and efficiency.

Motivation

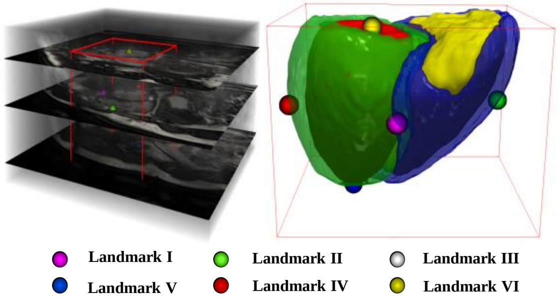

Fig. 1: Bounding box and landmarks in a CMR volumetric image. The left figure shows the location of the bounding box and six landmarks on a volume, and the right figure is the corresponding cardiac segmentation with six landmarks and the bounding box.

- MT3D is a multi-task learning framework tailored for medical imaging.
- Detect both anatomical landmarks and bounding box from CMR volumes.
- Extends the SSD from single-task learning to multitask learning and from 2D to 3D.
- The proposed framework is fast and accurate.

Method

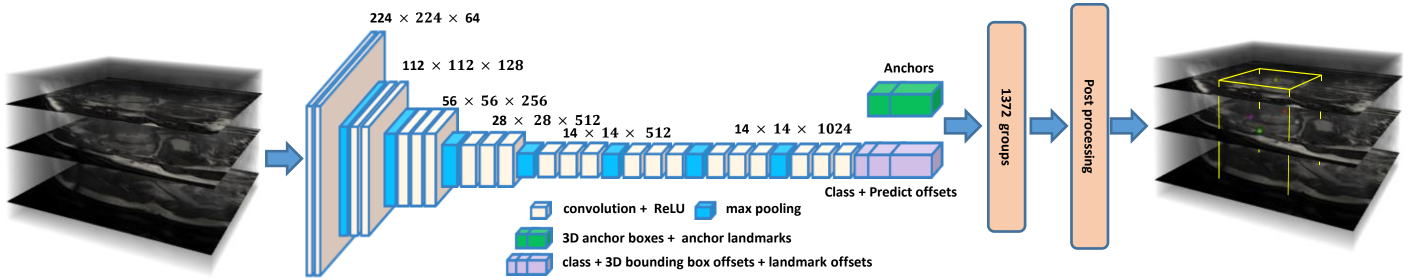

Fig. 2:The architecture of MT3D. MT3D adopts VGG16 as backbone network, which takes a whole volumetric CMR image as input, gradually downsamples the feature maps, and predicts the confidences and the offset for the anchor groups (bounding box and six landmarks).

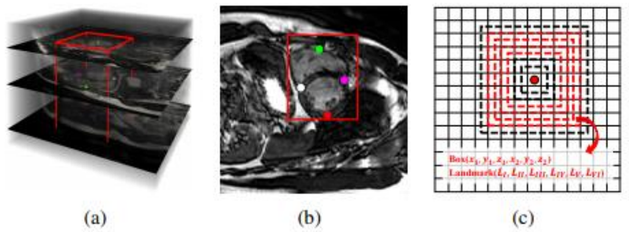
Box($x_1, y_1, z_1, x_2, y_2, z_2$)
Landmark($L_I, L_{II}, L_{III}, L_{IV}, L_V, L_{VI}$)
(a)
(b)
(c)

Fig. 3: MT3D takes the entire volume as input and predicts six landmarks within a bounding box.

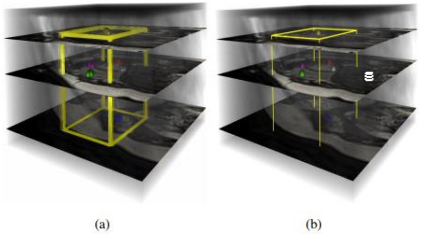

Fig. 4: A post processing example from validation set.

Bounding box detection

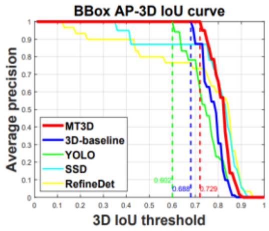

Fig. 5: The variations of AP with various 3D IoU thresholds

TABLE I: Comparisons of detection performance and running speed between different methods

Approaches	B-wall	B-centroid	VPS
YOLO V2 [24]	4.39	4.61	2.0
SSD [7]	3.93	6.24	1.6
RefineDet [22]	8.60	12.08	1.4
3D-baseline	4.02	4.42	99
MT3D	**3.11**	**4.38**	**128**

Landmark localization

TABLE II: Comparisons of landmark location accuracy and running speed. In ”L-Average” column, we report both NMS and average location result (NMS/average location) for 3Dbaseline and MT3D

approach	L-I	L-II	L-III	L-IV	L-VI	L-V	L-Average	VPS
human	7.55	21.78	8.77	11.51	5.84	10.12	10.93	-
Multi-SSLLN	5.64	13.75	**5.43**	10.74	6.35	9.42	8.56	31
single-SSLLN	10.98	**8.06**	6.44	6.53	**5.49**	**6.00**	7.25	31
3D-baseline	5.89	11.13	6.94	6.74	8.86	7.95	8.11/7.92	99
MT3D	**5.02**	8.38	7.36	**6.32**	6.87	6.58	6.91/**6.75**	**128**

Segmentation

TABLE III: The segmentation comparisons of before and after the Prediction bounding box is used to crop the CMR image.

	Dice Index		Hausdorff Dist. (mm)	
	full	cropped	full	corpped
LVC	0.9385	**0.9491**	3.6793	**3.2875**
LVW	0.8288	**0.8416**	4.9195	**4.2860**
RVC	0.87551	**0.8808**	**7.0745**	7.1443
RVW	0.6231	**0.6384**	11.5921	**8.4534**
VPS	31	**105**	31	**105**