



Weizeng Lu¹, Xi Jia², Wei Chen², Wei Chen², Nicolo Savioli ³, Antonio de Marvao³, Linlin Shen^{1*}, , Declan O'Regan³ and Jinming Duan ²

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

²School of Computer Science, University of Birmingham, United Kingdom
³MRC London Institute of Medical Sciences, Imperial College London, United Kingdom





Introduction



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.









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).





Anchor setting



Fig. 3: MT3D takes the entire volume as input and predicts six landmarks within a bounding box. (a) illustrates three slices and the ground truth group of the heart. (b) is the midventricular slice with the ground cross-section box and four landmarks on this slice. (c) represents the last layer feature map of the backbone, anchor landmarks and the cross-section boxes of 3D anchor boxes.





Post processing



Fig. 4: A post processing example from validation set. (a) shows the direct output of MT3D, i.e. the candidate bounding boxes and landmarks. (b) is the final output on the volume. We select the final bounding box using NMS algorithm, and the final six landmarks by averaging all candidate landmarks.





Detection



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



Fig. 6: Exemplary results using 2D methods. While 2D detectors are accurate for mid ventricular slices (top), they produce incorrect detections in basal/apical slices (bottom).





Detection



Fig. 7: The distribution of bounding box wall distance and centroid distance.

TABLE I: Comparisons of detection performance and running speed between different methods, where Bwall and B-centroid respectively represent the wall distance and centroid distance.

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. The LVC, LVW, RVC, and RVM denote the left ventricular cavity, left ventricular wall, right ventricular cavity, and right ventricular wall, respectively. And volumes per second (VPS) indicates the running speed.

	Dice	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		





Thank you