

Supplementary Material

Qinji Yu^{1,2}, Yirui Wang², Ke Yan^{2,3}, Le Lu², Na Shen⁴, Xianghua Ye⁵,
Xiaowei Ding¹, and Dakai Jin²

¹ Shanghai Jiao Tong University, Shanghai, China

² DAMO Academy, Alibaba Group

³ Hupan Lab, 310023, Hangzhou, China

⁴ Zhongshan Hospital Fudan University, Shanghai, China

⁵ The First Affiliated Hospital Zhejiang University, Hangzhou, China
dingxiaowei@sjtu.edu.cn, dakai.jin@alibaba-inc.com

Table 1: Statistics of 5 LN detection datasets, where 3 of them are used as the internal data for the model development and internal testing and the rest 2 are used as independent external testing set. NIH-LN is a public dataset, and other datasets are in-house datasets collected from 4 different clinical centers. HN, Eso and Mul. represent head & neck cancer, esophageal cancer and multiple types of diseases, respectively.

Dataset	#Patient	#LN	Avg. Res. (mm)	Body Parts	Setting
NIH-LN	89	1956	(0.82, 0.82, 2.0)	chest & abdomen	internal
Center1-HN	256	1890	(0.46, 0.46, 4.0)	head & neck	
Center2-Eso	91	857	(0.70, 0.70, 4.9)	chest	
Center3-Mul	184	2131	(0.76, 0.76, 2.0)	chest & abdomen	external
Center4-HN	50	418	(0.48, 0.48, 1.2)	head & neck	
Total	670	7252	-	-	-

Table 2: Detailed performance of our method across *different size* lymph nodes (Recall@FPs=[0.5, 1, 2, 4]).

Ln size	Recall@FPs=[0.5, 1, 2, 4].				
	@0.5	@1	@2	@4	Avg.
All size	26.00	33.86	44.88	54.16	39.73
≥ 5mm	39.70	49.94	60.75	69.38	54.94
≥ 7mm	54.27	65.18	74.08	79.71	68.31
≥ 10mm	66.39	76.12	78.94	83.24	76.17

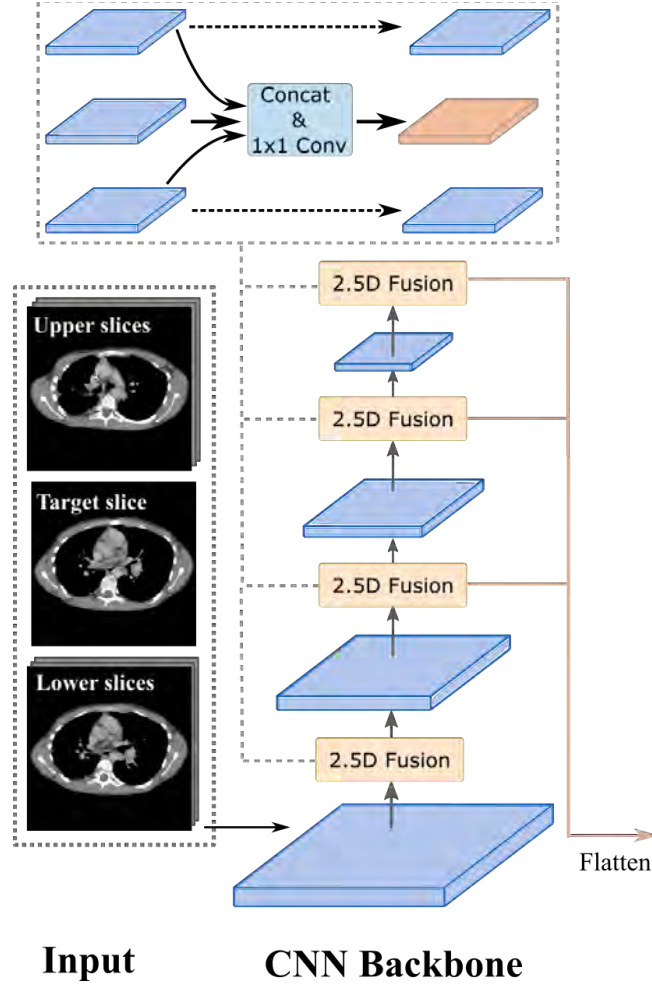


Fig. 1: Details of our 2.5D backbone and the multi-scale 2.5D feature fusion scheme. Specifically, to detect LN in a target CT slice, we extract four upper and four lower slices from the original CT scan to serve as the 3D context of the central target slice. Then, these total nine consecutive CT slices are grouped into three sets of 3-channel images. Each set is processed by the shared CNN backbone independently, and the three sets are then fused by concatenation and 1×1 conv. After that, the feature map of the original central target slice is replaced by the fused feature map, while the upper and lower sets' feature maps remain unchanged.