Supplementary Material: DKMA-ULD: Domain-Knowledge augmented Multi-head Attention based Robust Universal Lesion Detection

Manu Sheoran*
manu.sheoran@tcs.com
Meghal Dani*
dani.meghal@tcs.com
Monika Sharma
monika.sharma1@tcs.com
Lovekesh Vig
lovekesh.vig@tcs.com

TCS Research
New Delhi, India

1 Ablation for Feature Extraction Backbone

We use 5 multiple intensity images having 3 slices/channels each for a given patient in DeepLesion [4] dataset. These images are passed as input to the shared convolutional feature extractor with feature pyramid network (FPN) [3]. To determine the best feature extractor network that can learn the most relevant features for 3D CT-scans, we performed experiments with different ResNet [2] variants. We found that ResNeXt-152 performs the best, as it is evident in Table 1. Therefore, in our proposed *DKMA-ULD* method, we utilize the ResNeXt-152 backbone for feature extraction with self-supervised weights using BYOL technique [1] and achieve state-of-the-art average sensitivity of 87.16% on the test-set of the DeepLesion dataset.

Model	Backbone	FP@0.5	FP@1.0	FP@2.0	FP@4.0	Average
DKMA-ULD	**x101**	75.09	83.88	89.28	92.83	85.27
DKMA-ULD + BYOL	**x101**	76.07	84.31	89.44	92.94	85.69
DKMA-ULD	**x152**	78.10	85.26	90.48	93.48	86.88
DKMA-ULD + BYOL	**x152**	**78.75**	**85.95**	**90.48**	**93.48**	**87.16**

Table 1: Sensitivity(%) for DKMA-ULD using different backbone networks and weight initialization via self-supervised learning (SSL), at different false-positives (FP) per sub-volume on the volumetric test-set of DeepLesion [4] dataset.

* Equal contribution in paper

2 Organ-wise sensitivity

DeepLesion dataset [4] consists of approx. $32K$ annotated lesions across 8 different organs of the body. It is the largest dataset available right now which contains lesions in a variety of organs and hence, the best candidate for developing a universal lesion detection network. Our proposed method *DKMA-ULD* uses multiple-HU windows and lesion-specific custom anchors with a novel multi-head self attention-based feature fusion module. The inclusion of domain-knowledge specific features in the proposed *DKMA-ULD* has led to an improved overall and organ-wise performance as evident in Table 2.

Organ Type	FP@0.5	FP@1.0	FP@2.0	FP@0.5	Average
Bone	65.74	79.63	84.26	87.03	79.16
Lung	89.45	93.27	95.72	96.45	93.72
Mediastinum	84.34	90.37	93.27	95.35	90.83
Liver	83.00	89.00	93.42	95.57	90.24
Kidney	77.58	82.75	91.37	93.10	86.20
Abdomen	67.88	78.96	85.52	90.12	80.62
Pelvis	76.83	83.41	88.29	91.71	85.06
Soft Tissue	62.75	70.96	78.59	84.16	74.11

Table 2: Organ-wise sensitivity in % (at different FP and average) for *DKMA-ULD* having BYOL initialized ResNeXt-152 backbone.

Method	BN	LNG	MDT	LVR	KDY	ABM	PLS	ST
MULAN (w/o tags) [5]	77.31	89.86	88.37	88.75	76.83	79.69	83.84	71.80
DKMA-ULD(ours w/o byol)	**79.16**	**94.30**	**89.32**	**89.64**	**84.37**	**80.81**	**85.97**	**76.31**

Table 3: Organ-wise average sensitivity (%) comparison (over $FP = \{0.5, 1, 2, 4\}$) of base MULAN [5] model w/o tags and our proposed base *DKMA-ULD* model (without byol self-supervision) on test-set of DeepLesion [4] dataset.

Next, we provide detailed comparison of our proposed DKMA-ULD network with MULAN [5]. Since, the best trained model of MULAN with tags is not available publicly, we use the official released base model of MULAN [5] for computing organ-wise sensitivity values on DeepLesion test-set for comparison. For a fair comparison, we use the base model of our proposed DKMA-ULD without self-supervision in Table 3. It is clearly visible that our proposed method DKMA-ULD outperforms MULAN in lesion detection across all organs. Further, We present a qualitative comparison of our proposed DKMA-ULD with MULAN [5] in Figure 1 and demonstrate that the detection of false positives is substantially reduced using domain-knowledge.

References

[1] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733*, 2020.

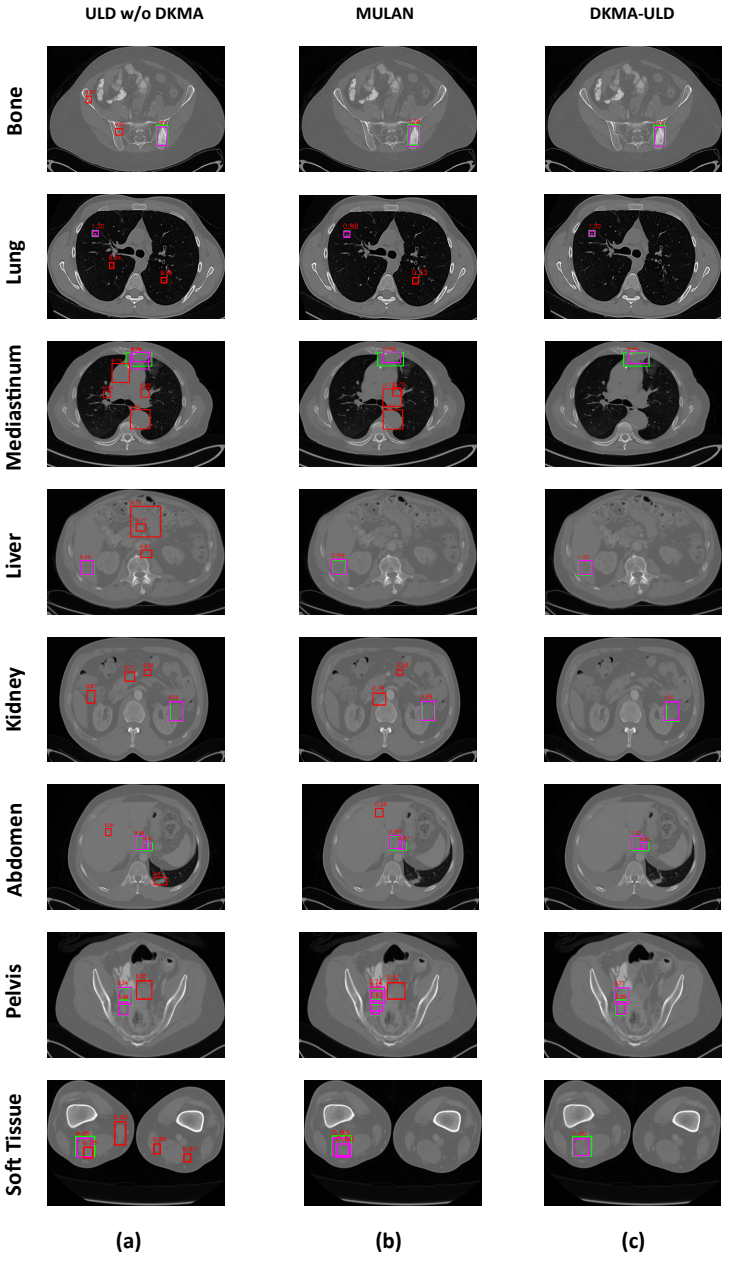

Figure 1: Qualitative comparison of *DKMA-ULD* and MULAN [5] (at FP =2) on CT-scans of different body regions. The green, magenta, and red color boxes represent ground-truth, true-positive (TP), and false-positive (FP) lesion detection, respectively. Please note that ULD w/o DKMA represents when 3 slices with only one HU window ($[1024, 4096]$), default anchors, and without convolution augmented multi-head attention feature fusion are used. We can observe that after incorporating domain knowledge in the form of multi-intensity CT slices, custom anchors, and multi-head attention (i.e., *DKMA-ULD*), the number of FP reduced drastically resulting in improved lesion detection performance as compared to MULAN.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[3] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[4] Ke Yan et al. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J. Med. Imaging*, 5(3):036501, 2018.

[5] Ke Yan et al. Mulan: multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation. In *MICCAI*, pages 194–202. Springer, 2019.