Appendix: CancerUniT

Abstract. This document contains the Supplementary Materials for the ICCV 2023 paper ”CancerUniT: Towards a Single Unified Model for Effective Detection, Segmentation, and Diagnosis of Eight Major Cancers Using a Large Collection of CT Scans”. It covers the model generalizability to public dataset, (§A), model instantiation details (§B), semantic segmentation results of full spectrum tumors (§C), and the qualitative results (§E).

A. Generalizability to Public Dataset

Our method aims at holistically modeling the multiple cancer screening problem versus non-cancer. However, to the best of our knowledge, no public dataset is suitable for such problems. Nevertheless, our trained model generalizes well on three public single-tumor datasets including MSD pancreas, liver and lung dataset, as shown in Table. A. It is worth noting that our model inference directly without extra training, whereas the 3 single-nnUNets is trained on the MSD dataset with domain knowledge. To be specific, the experiment of 3 single-nnUNet is conducted with 5-fold cross-validation, and our UniT is tested on the same validation set.

Despite not having any prior knowledge of the data distribution, our proposed UniT model effectively suppresses the single-tumor expert model, achieving an average tumor detection sensitivity improvement of 3.1%. Our results demonstrate the efficacy of our proposed method for addressing the tumor detection problem without the need for a specific dataset. The ability to generalize well on public datasets and suppress the single-tumor expert model underscores the potential of our approach to be used as a practical solution for universal cancer screening and diagnosis.

Table 8. Generalizability to 3 Public MSD dataset [2]. Average detection sensitivity is reported. Our model inference directly, whereas 3 single-nnUNets are trained on the MSD dataset.

	Pancreatic	Liver	Lung	Avg	Speed	Param
	tumor	tumor	tumor			
single-nnUNet (trained)	88%	97%	90.5%	91.8%	66s	92.34M
Ours (test)	94.7%	93.1%	97%	94.9%	42s	30.87M

B. Model Instantiation Details

In our UniT, the hidden dimension of query is set to 32, such that the detection query $\mathbf{A}^j \in \mathbb{R}^{4\times32}$, the diagnosis query $\mathbf{B}^j \in \mathbb{R}^{10\times32}$, the shared query $\mathbf{S}^j \in \mathbb{R}^{12\times32}$. We adopt nnUNet [26] as the backbone to extract multi-scale features $\mathbf{F} = [\mathbf{F}^1, \mathbf{F}^2, \mathbf{F}^3, \mathbf{F}^4]$. Note, $\mathbf{F}^j \in \mathbb{R}^{d\times(D\times H\times W)}$ is flatten and projected from intermediate spatial feature $\hat{\mathbf{F}}^j \in \mathbb{R}^{C\times D\times H\times W}$. In specific, $\mathbf{F}^1 \in \mathbb{R}^{32\times(48\times192\times192)}$, $\mathbf{F}^2 \in \mathbb{R}^{32\times(48\times96\times96)}$, $\mathbf{F}^3 \in \mathbb{R}^{32\times(24\times48\times48)}$, and $\mathbf{F}^4 \in \mathbb{R}^{32\times(12\times24\times24)}$. The total number of Transformer layer is set to 3, each of which contains a multi-head cross-attention, a multi-head self-attention, and a feed-forward network. Note, in the inference stage, the tumor segmentation maps are extract to generate the tumor instances with class labels, where those tumor instances with less than 200 voxels are discarded.

C. Semantic Segmentation Results of Full Spectrum Tumors

We conducted an evaluation of our model's performance on the semantic segmentation of full spectrum tumors, which is a challenging task that involves the segmentation of multiple tumor subtypes within an organ. The quality of the multi-class tumor segmentation was assessed using the multi-class Dice score, where each subtype of the tumor was treated as an independent semantic class.

Our model outperformed the segmentation baselines and achieved the highest average segmentation Dice score, as demonstrated in Table 9. Notably, our model was not compared with detection models such as DeepLesion and LENS, as these models are not designed for semantic segmentation tasks.

Our findings suggest that enhancing the query hierarchy in our model can improve the semantic segmentation of full spectrum tumors. This observation is in line with our assumption that our query-based Transformer model can more effectively explore the similarity between intra-organ tumor subtypes, leading to improved segmentation performance. Overall, our evaluation provides evidence that our proposed model can effectively address the challenges of multi-class tumor segmentation in the context of full spectrum tumors.

D. Universal Cancer Screening: CT *vs* Blood Test.

Blood test is now one of the most attractive tools for non-invasive multi-organ cancer screening [13, 30, 28]. CT scanning had been considered historically for the same task, but was limited by its insufficient sensitivity and specificity [1]. AI reading in CT as an alternative opportunistic screening tool, our approach also has strong clinical potential for cancer detection screening. The advantage of CT is that this protocol is already an indispensable diagnostic imaging for cancer, but a positive blood test result requires further examinations for confirmation. With our model, clinicians have direct visual analyses of the detected cancer sites and misdetections of cancer can be largely reduced. No additional cost is needed under the opportunistic CT screening protocol whereas a single blood test can usually take $\sim$ 1000 US

Table 9. Voxel-level semantic segmentation results of full spectrum tumors. The Dice coefficient is reported. Note: the Dice values are calculated in a semantic manner, e.g., the HCC voxel is correctly segmented as the HCC subtype (not other liver tumor subtypes or other tumor types) by the semantic segmentation.

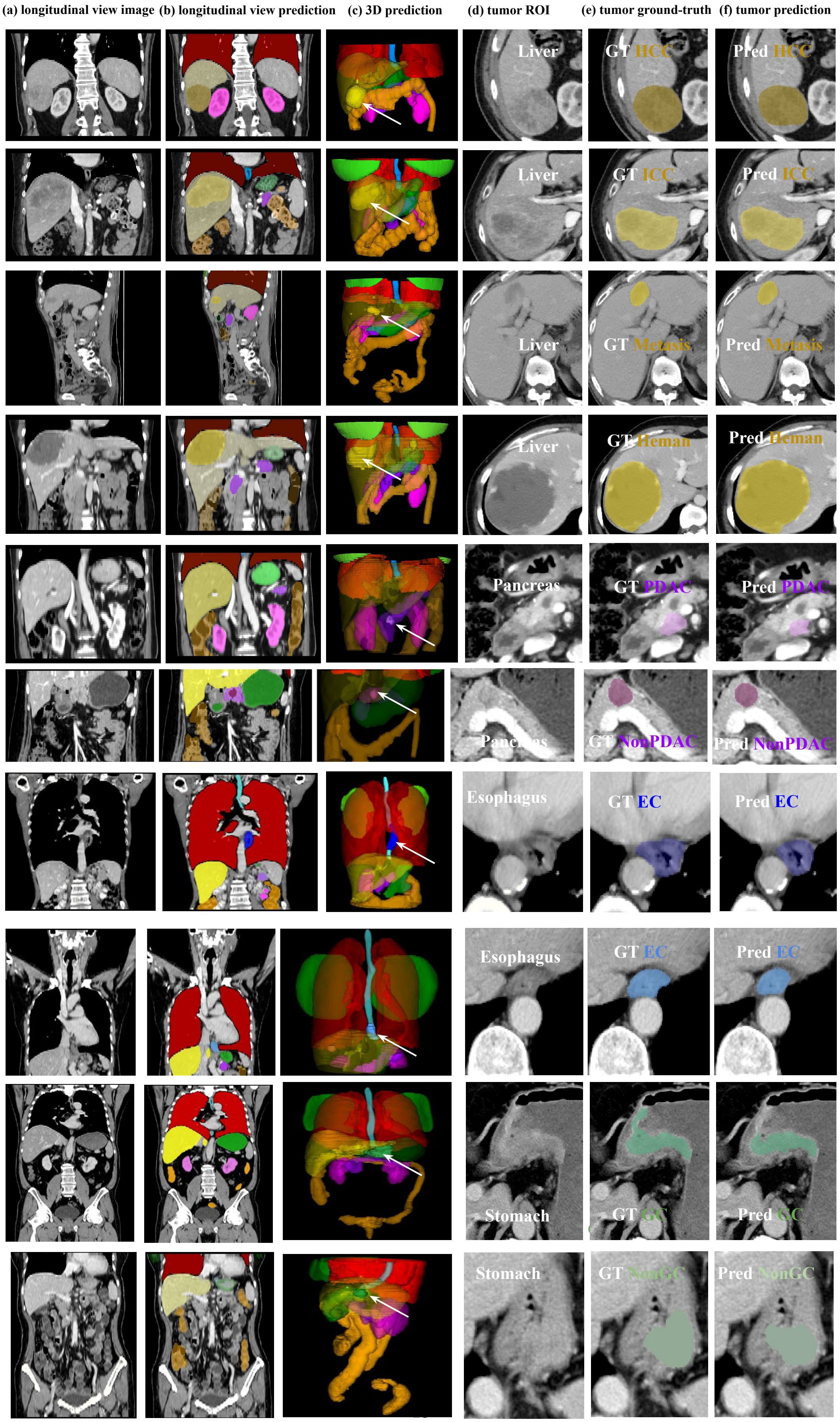
Model	Pancreas			Eso			Stomach			Liver					Average
	PDAC	nonPDAC	avg	EC	nonEC	avg	GC	nonGC	avg	HCC	ICC	Meta	Heman	avg	
8-nnUNet ensemble	0.750	0.525	0.638	0.770	0.433	0.602	0.441	0.099	0.270	0.489	0.552	0.296	0.784	0.530	0.510
nnUNet [26]	0.758	0.534	0.646	0.739	0.207	0.473	0.453	0.068	0.261	0.410	0.481	0.306	0.739	0.484	0.466
TransUNet [7]	0.749	0.553	0.651	0.744	0.321	0.533	0.473	0.128	0.301	0.411	0.503	0.353	0.717	0.496	0.495
Ours	0.728	0.560	0.644	0.738	0.457	0.597	0.389	0.187	0.288	0.368	0.666	0.305	0.773	0.528	0.514

dollars.

For relative performance comparison to CancerSeek [13], i.e., cancer vs. normal, our method has higher sensitivity levels in detecting six out of seven types of cancers: approximately for stomach (+18%), pancreas (+24%), esophagus (+26%), colorectum (+35%), lung (+34%), and breast (+57%). Our averaged patient-level cancer detection sensitivity is 94% versus 70% in [13]. The test specificity for normal cases in venous CT is 100% (blood test>99%). We acknowledge that the results of the representative blood test [13] and ours may not directly comparable since different test data are used. Nevertheless, the rough comparison shows the high accuracy of CT+AI solution, and thus may re-open doors for multi-cancer screening by CT [1].

E. Qualitative Results

We provide more qualitative results of full spectrum tumors in the test set being segmented and diagnosed by our method as shown in Fig. 4. The results demonstrate that our method can not only segment the tumor region well but also predict the class of tumor subtype correctly.



(a) longitudinal view image (b) longitudinal view prediction (c) 3D prediction (d) tumor ROI (e) tumor ground-truth (f) tumor prediction

Figure 4. Qualitative results of full spectrum tumors in the test set being segmented and diagnosed by our method (best viewed in color).