

VISTA3D: A Unified Segmentation Foundation Model For 3D Medical Imaging

—Supplementary Material—

Supplementary Material

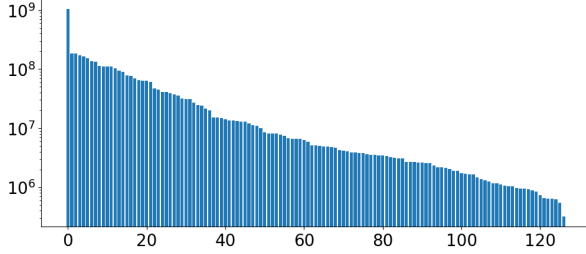


Figure 1. Distribution of annotated voxels in the training set (X-axis: class index, Y-axis: number of annotated voxels per class).

1. Dataset Details

Table 1 lists more details about our curated dataset. Fig. 1 shows the number of annotated voxels according to the corresponding task classes. Spatial resolutions range from $0.45 \times 0.45 \times 0.45$ to $1.50 \times 1.50 \times 7.50$ (median: $0.88 \times 0.88 \times 1.50$) mm^3 .

Global and local index for the partial label Those datasets have different number of classes and indexes in their manual labels (e.g. Pancreas in MSD07 has index 1 but 10 in TotalSegmentatorV2). We curated a global index of 127 integers and mapped all local indexes in each individual dataset to this global index. We also curated a **label set** list for each dataset, containing the class index that will be used within this dataset. We included as much dataset with a commercial license as possible for the development of this method.

Dataset ID	Type	# cases used
TCIA Pancreas CT [12]	Abdominal CT organs	80
AbdomenCT-1K [9]	Abdominal CT organs	1,050
AMOS22 [7]	Abdominal CT organs	300
MSD Task 3,6,7,8,9,10 [1]	Various lesions	945
CT-ORG [11]	Lung, bones, liver, kidneys, bladder	136
TotalSegmentator [4]	Many anatomic structures	1,228
CRML-CT [15]	Liver, colorectal liver metastases	197
VerSe [14]	Vertebral labelling	374
AeroPath [16]	Airways and lungs	27
Bone lesion (in-house)	bones	296
LIDC-IDRI [2]	Unannotated, lung cancer screening thoracic	470
COVID-19 [5]	Unannotated, chest	524
TCIA Colonography [8]	Unannotated, abdomen	1,440
StonyBrook COVID19 CT [13]	Unannotated, chest	1,274
NLST [17]	Unannotated, chest	3,113

Table 1. Summary of datasets used for model training.

2. Computational Details

Training Requirements The model is trained on 64 32GB NVIDIA V100 GPUs with around 20,000 total GPU hours. The prompt number (object class) in a single training iteration is 36 for automatic branch, and 4 for point branch. The

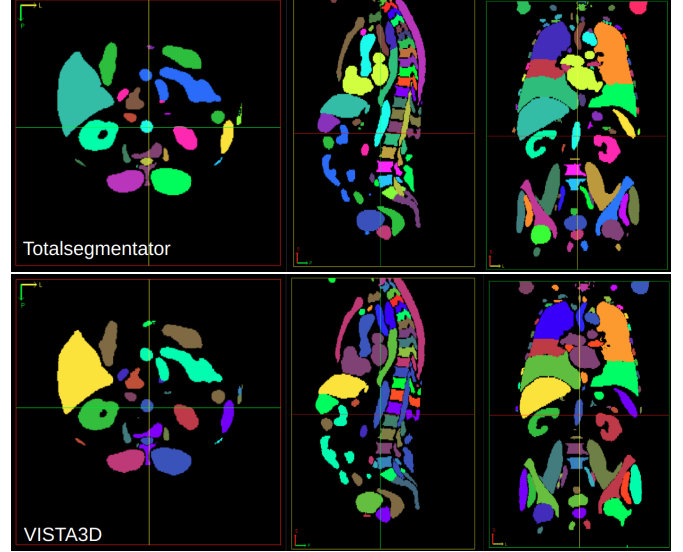


Figure 2. An example of whole class segmentation on a typical sized human CT scan. Running on a lower end machine with 12GB GPU. The runtime for VISTA3D is 1m43s and 2m41s for TotalSegmentator.

model can be trained with 16GB memory GPUs or even lower by reducing the prompt number in each iteration, at the cost of longer training time if number of classes is large.

Inference Requirements The inference GPU memory requirements also depend on the prompt number and image size. Since the model is based on sliding window of size $128 \times 128 \times 128$, the GPU memory requirements can be optimized to be stable and less dependent on image size. We used a sliding window inferer with adaptive memory control to switch between CPU and GPU to avoid the out-of-memory issue. We benchmarked the runtime on a 16GB V100 GPU in the main paper. TotalSegmentator uses 5 sub-task models for different class groups and thus can be slower. Here we also performed inference on a lower-end environment with 12GB memory GPU and 32GB memory CPU. The results of a typical CT scan (a MSD task03 test scan [1], size $308 \times 260 \times 453$ after resampling) are shown in Fig. 2. The runtime for VISTA3D is 1m43s and 2m41s for the TotalSegmentator. For interactive segmentation, the single-click point inference run-time is 3.2s on the same 12GB GPU machine. Two examples are shown in Fig. 3.

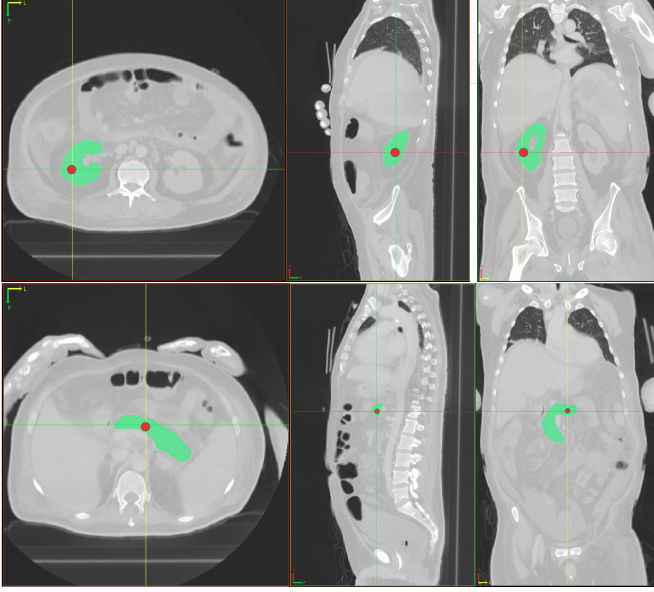


Figure 3. An example of using single point click for organ segmentation. Runtime on a machine with 12GB GPU and 32GB CPU is constant 3.2 seconds, regardless of image size or organ size.

3. Additional training details

3.1. Stage1-Interactive branch training

The algorithm is shown in Alg. 1. The point sampler S works as a data augmenter, with 50% probability to sample points directly from y to get the point p and binary groundtruth mask y_{gt} as a training pair, while another 50% will be used with the following augmentations: a) random sample points from supervoxel and form a zero-shot training pair. b) random add or subtract a supervoxel mask that satisfies a certain size and position criterion to y , this is used to force the model to be able to edit supported class mask. Meanwhile, when the subtraction or addition size exceeds a certain limit, the generated training pair will also be used as zero-shot pairs with the zero-shot embedding. We use $max_{iter} = 5$ for the training.

3.2. Stage3-Automatic branch training

For each patch, we randomly sample the existing class indexes c from its manual label or pseudo-label and obtain the corresponding binary mask y_{gt} or y_{gt}^p . The algorithm is shown in Alg. 2. Unlike traditional segmentation models that do softmax on multichannel output, our automatic segmentation is based on promptable binary segmentation, thus prone to produce false positives. We mitigate this issue by sampling the background prompts from $label_set - y.unique()$ or $label_set - y^p.unique()$ and train the model to produce zero output when responding to the prompt. So in each iteration, a 128 cubic image patch is

Algorithm 1 Interactive branch training

Require: VISTA interactive branch model Φ , image patch x , image manual label y , image pseudo label y_p , supervoxel y_s .

Ensure: At least one of y or y_p are not None

```

 $S \leftarrow point\_sampler(y, y_s)$   $\triangleright$  Initialize point sampler based on manual label and supervoxel
 $S_p \leftarrow point\_sampler(y_p, y_s)$ 
 $p, y_{gt} \leftarrow S.sample()$   $\triangleright$  Sample point prompts  $p$  and segmentation mask  $y_{gt}$ 
 $p^p, y_{gt}^p \leftarrow S.sample()$ 
for  $i = 1$  to  $max\_iter$  do
     $loss \leftarrow LossFunction(\Phi(x, p), y_{gt})$ 
     $loss_p \leftarrow LossFunction(\Phi(x, p^p), y_{gt}^p)$ 
    update  $\Phi$  using  $loss + loss_p$ 
     $p = p \cup Sample(\Phi(x, p), y_{gt})$   $\triangleright$  Sample 1 point each from false positive and negative region
     $p^p = p^p \cup Sample(\Phi(x, p^p), y_{gt}^p)$ 
end for

```

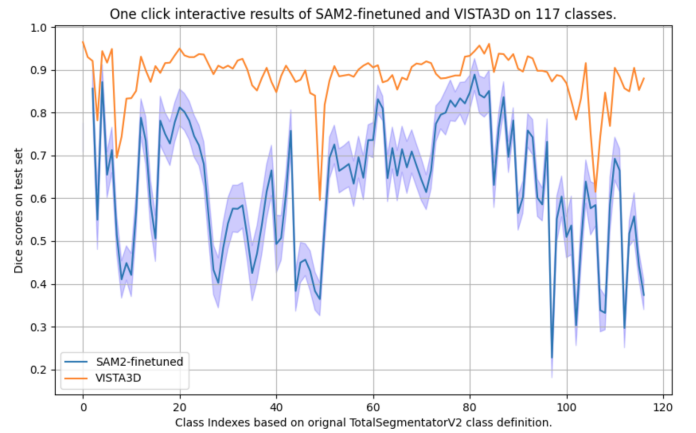


Figure 4. One-point interactive results for SAM2 and VISTA3D on TotalSegmentatorV2 test split. SAM2 is finetuned on the training split with SAM2’s official finetuning script. Click point selected at the center slice of each foreground.

the model input, and we sample a maximum of 32 class prompts using Alg. 2 and a maximum of four background prompts. All of those prompts are concatenated in the batch dimension.

3.3. SAM2 finetuning experiments

We performed a detailed fine-tuning experiment to see if the SAM2 object tracking-based algorithm can be applied to 3D medical image segmentation. We used the official SAM2 finetuning code¹ and finetuned on the TotalSegmentator [4] training set. Each axial slice is considered a video

¹<https://github.com/facebookresearch/sam2/tree/main/training>

MSD Task07	Pancreas	Pancreas Tumor
VISTA3D	0.802	0.603
SAM2-Finetuned	0.557	0.308

Table 2. Single click performance on MSD Task07 test set. SAM2 finetuned only on the training split.

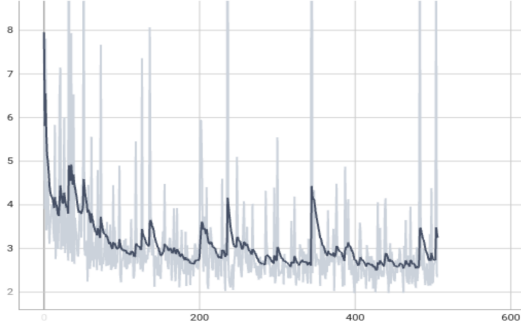


Figure 5. Loss curves for finetuning SAM2 on TotalSegmentator dataset.

frame. The model is trained on 8 80GB A100 GPU for 500 epochs until full convergence as shown in Fig. 5. However, the results as shown in Fig. 4 is disappointing. We also finetuned on MSD Task07 pancreas and pancreas tumor to reduce the class number. The results can be shown in Table. 2. The SAM2 method can track objects with simple shape and clear boundary very well, like femur bones, but failed to track complicated 3D shapes. Similar findings can be found in [6].

Algorithm 2 Automatic branch training

Require: VISTA automatic branch model Φ_a with encoder frozen, image patch x , image manual label y , image pseudo label y_p .

Ensure: At least one of y or y_p are not None

$c, y_{gt} \leftarrow y.unique().sample()$ ▷ Sample class prompts c and segmentation mask y_{gt}

$c^p, y_{gt}^p \leftarrow y_p.unique().sample()$

$loss \leftarrow LossFunction(\Phi_a(x, c), y_{gt})$

$loss_p \leftarrow LossFunction(\Phi_a(x, c^p), y_{gt}^p)$

update Φ_a using $loss + loss_p$

4. Additional Results

We provide additional VISTA3D results in this section. The baseline MedSAM [10] and SegVol [3] results are from their provided user interface and online hugging-face demo.

4.1. Qualitative Results

Editing examples We show an extreme example in Fig. 6, illustrating that VISTA3D supports detailed editing at pixel

level, while the bounding box prompt cannot perform any editing.

Hard Examples We show some hard classes like hepatic vessel and pancreas in Fig. 7 and Fig. 8. Those classes are included in VISTA3D’s and SegVol’s training sets. We randomly picked an abdominal scan from the MSD task09 spleen test split. This dataset does not contain annotations for pancreas or hepatic vessel, thus it can avoid groundtruth leakage and provide fair comparison.

Zero-shot interactive examples In Fig. 9, we show the interactive segmentation on a micro-CT mouse left lung. We can see that MedSAM has a major weakness of not being able to perform fine detailed editing, while SegVol’s response resolution is low. Fig. 10 shows other slices of the same mouse scan as Fig. 9. The figure shows a good point response on slices even far away from the clicks, illustrating the ability of 3D annotation and reducing annotation effort. We also provide additional illustrations of Mouse-CT dataset and our zero-shot results for left lung in Fig. 11.

Overfitting to common organs Due to the lack of diversity of 3D organs, the model can easily overfit to certain classes and remember the shapes, intensities, or locations. This is beneficial for achieving superior segmentation accuracy, but on the other hand, the model will ignore point clicks, even without providing any semantic information about the class. An example is shown in Fig. 12. We click a point outside of the kidney to segment the fluid around, and this should be zeroshot. SegVol directly segments the kidney and ignores the point. VISTA3D avoids this problem by using the zero-shot embedding and the novel model and recipe design. The area outside of organ is relatively easy; what if we want to forcefully segment a supported organ into sub-parts? We show an example in Fig. 13. If we click positive points on the liver, the model tends to ignore the points and directly segment the liver. Adding a zero-shot embedding will make the model follow the clicks much better.

4.2. Quantitative Results

We provide detailed Dice scores on all the classes of our test datasets. The result is shown in Table. 3.

5. Additional Discussions

The VISTA3D model design will naturally raise two questions, why not share decoder and why share encoder. If we share the encoder and decoder, then automatic and interactive will be trained together, which will 1) slow down the training. Interactive branch is much more memory intensive than automatic branch, and the supervoxels are only used for interactive training, thus, automatic branch can use a much larger batch size. Combine these two training will reduce automatic branch training iteration and its performance. 2) There are internal conflicts between zero-shot and automatic segmentation, our pilot study showed worse

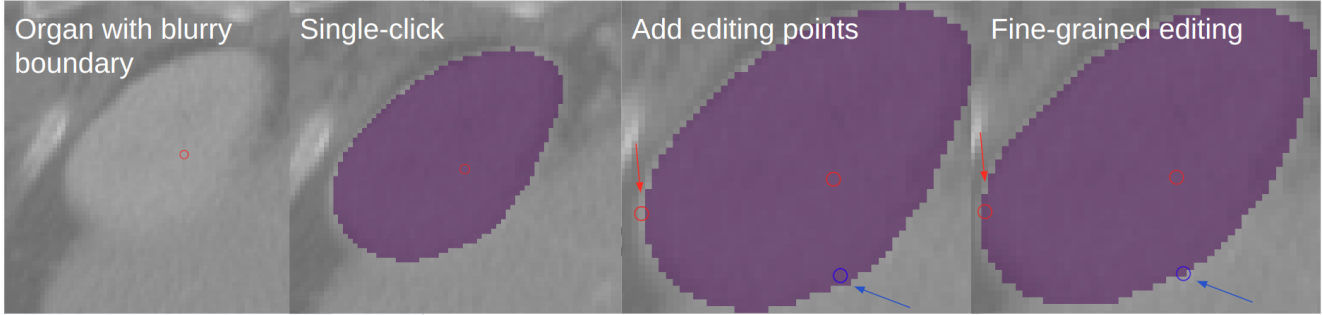


Figure 6. Fine-grained editing on blurry boundary. Red is positive point and blue is the negative point. This is an extreme example to show that VISTA3D can edit one-pixel wide boundaries. The addition or removal area depends on the model’s understanding of boundaries, and the edited area by a single click could be much larger.

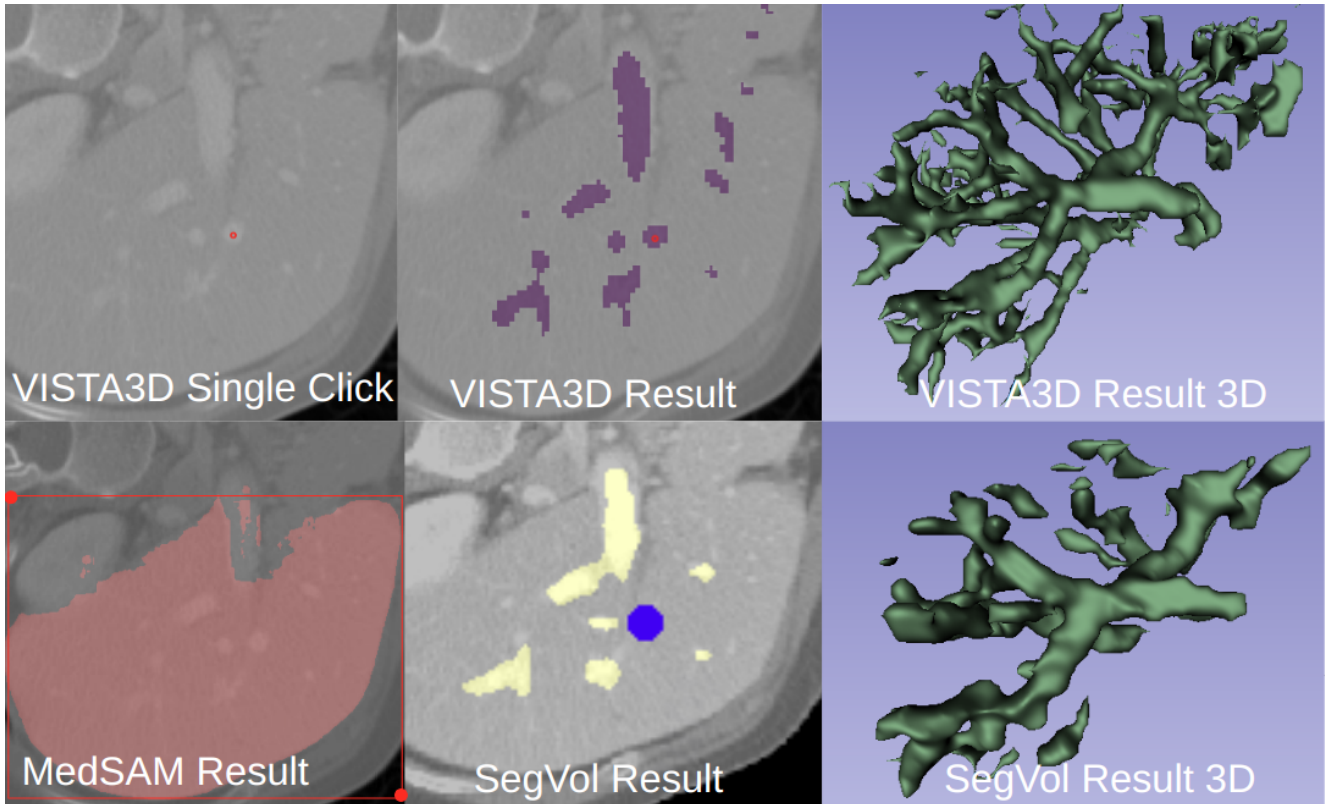


Figure 7. Single positive point for hepatic vessel segmentation (example from MSD09 spleen held out test set, no hepatic vessel groundtruth to avoid groundtruth leakage). SegVol demo uses blue dot while VISTA3D demo uses red dot to represent positive clicks. VISTA3D achieved much better results in details.

results and our auto-branch is not able to reach state-of-the-art results once trained together with interactive branch. Sharing encoder has two purposes, 1) we support interactive editing over automatic results, the shared encoder could reduce inference computation cost. 2) The interactive branch can be trained with a much broader range of data, thus the encoder can extract more generalizable features and help with the generalizability of automatic segmentation.

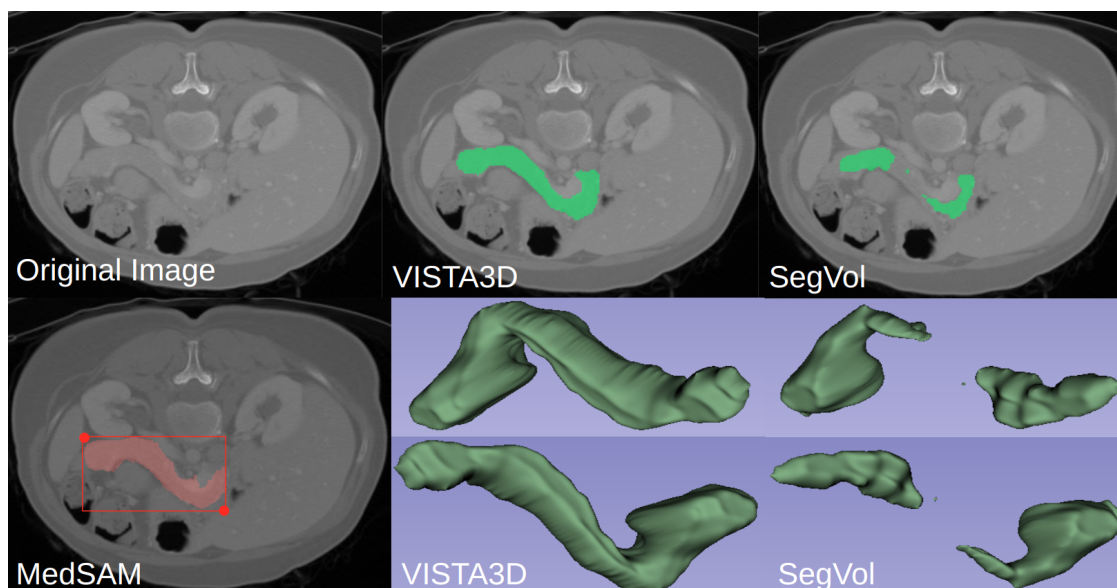


Figure 8. Automatic (semantic) segmentation for pancreas (example from MSD09 spleen held out test set, no pancreas groundtruth). VISTA3D achieves much better results in details and segmentation completeness.

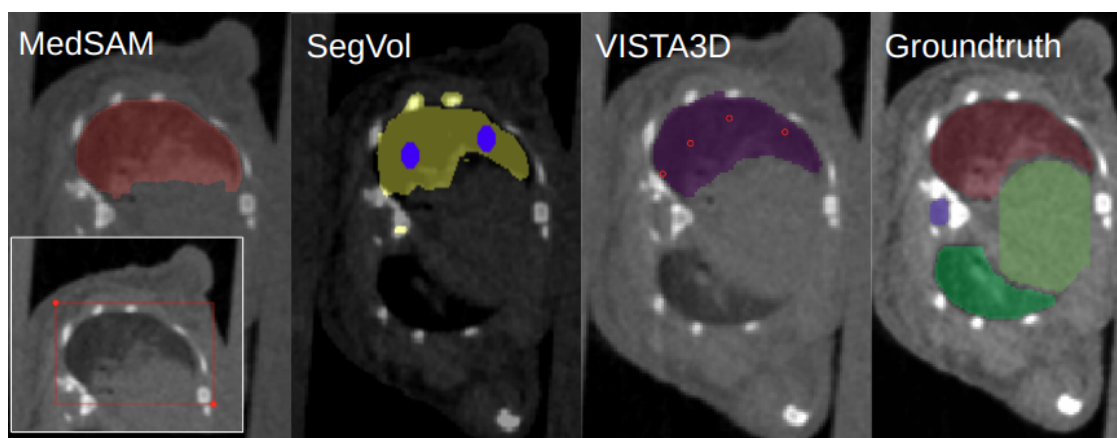


Figure 9. Interactive segmentation on micro-CT mouse left lung. Baseline results from MedSAM local user interface and SegVol demo.

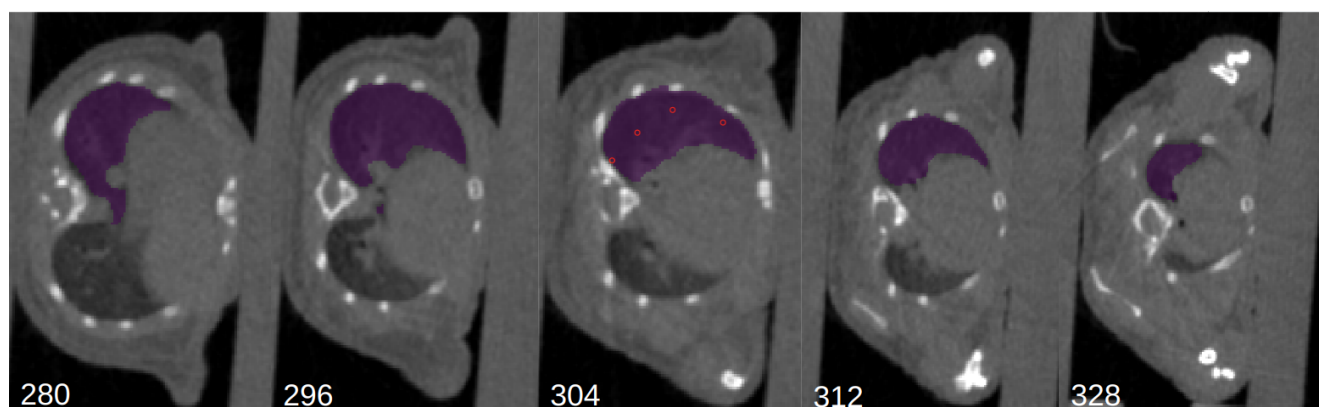


Figure 10. 3D point response on far away slices. The point click is on slice 304 (same as Fig. 9), but the segmentation on slices 280, 296, 312, and 328 all showed good results, showing the potential of reducing annotation effort in 3D space.

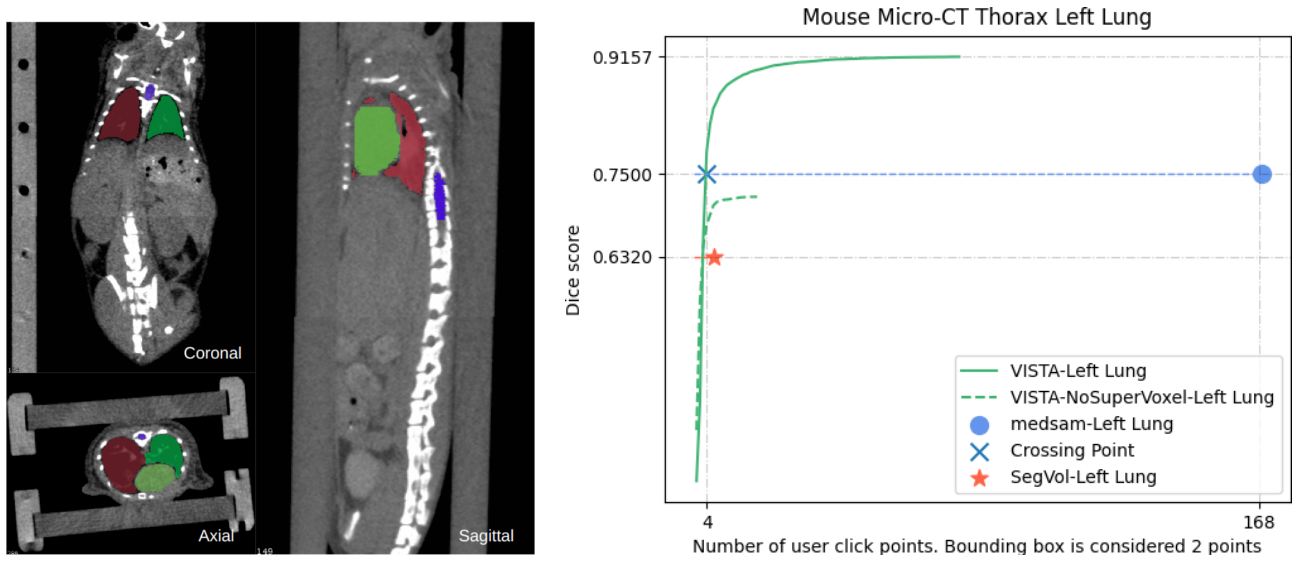


Figure 11. The mouse micro-CT example and the mouse left lung zero-shot performances. Even "left lung" is in the supported class, the huge structural difference between human and mice will fail any automatic segmentation model trained on human anatomy.

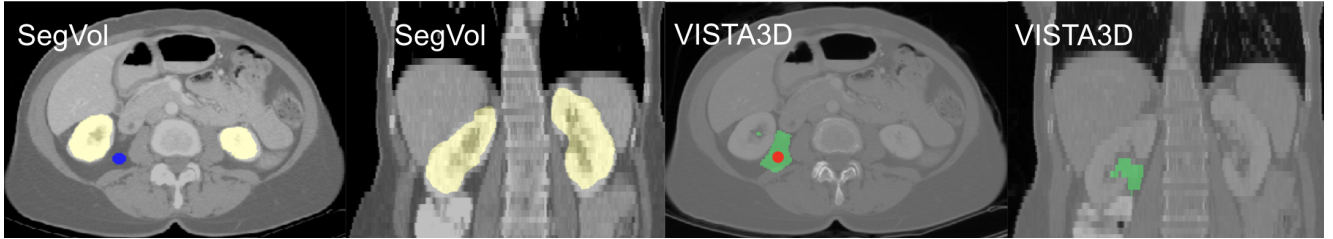


Figure 12. The overfitting problem with common organs. Due to the lack of diversity of 3D organs, the model can easily overfit to certain classes and ignore point clicks, even without providing any semantic information about the class. We click a positive point outside of kidney to segment the fluid around, and this should be zeroshot. SegVol directly overfits to segment kidney and ignores the points. VISTA3D avoided this problem by using the zero-shot embedding and the novel model and recipe design.

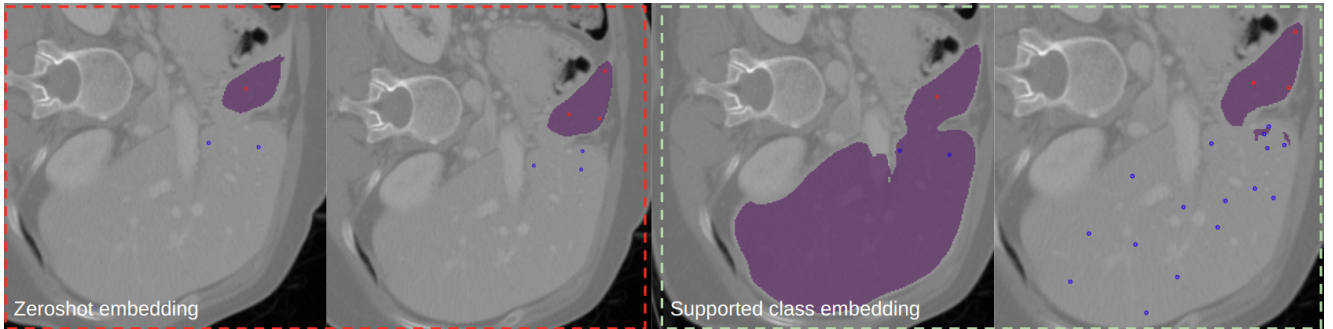


Figure 13. Use points to forcefully separate liver into substructures. We can see that VISTA3D with zero-shot embedding responds much better to the clicks. However, if the model uses supported class embedding, the model is reluctant to respond to negative points for liver segmentation.

Table 3. Dice score of all the classes on the test datasets.

	Auto-3dSeg	nnUNet	TotalSeg-mentator	VISTA3D auto	VISTA3D point	VISTA3D auto + point
MSD03 Hepatic Tumor						
liver	0.943	0.947	0.942	0.959	0.874	0.961
hepatic tumor	0.616	0.617	-	0.588	0.701	0.687
MSD06 Lung Tumor						
lung tumor	0.562	0.554	-	0.614	0.682	0.719
MSD07 Pancreatic Tumor						
pancreas	0.785	0.789	0.775	0.819	0.802	0.840
pancreatic tumor	0.485	0.488	-	0.324	0.603	0.638
MSD08 Hepatic Tumor						
hepatic vessel	0.627	0.584	-	0.553	0.582	0.670
hepatic tumor	0.683	0.659	-	0.682	0.733	0.757
MSD09 Spleen						
spleen	0.965	0.967	0.935	0.952	0.938	0.954
MSD10 Colon Tumor						
colon cancer primaries	0.475	0.473	-	0.439	0.609	0.633
AeroPath						
lung	0.982	0.974	0.957	-	-	-
airway	0.896	0.899	-	0.852	0.819	0.867
Bone Lesions						
bone lesions	0.343	0.396	-	0.491	0.536	0.585
BTCV-Abdomen						
spleen	0.954	0.962	0.951	0.944	0.950	0.955
right kidney	0.936	0.951	0.941	0.943	0.937	0.945
left kidney	0.942	0.932	0.944	0.942	0.938	0.946
gallbladder	0.663	0.771	0.739	0.794	0.792	0.807
esophagus	0.740	0.740	0.793	0.779	0.799	0.821
liver	0.964	0.961	0.970	0.967	0.715	0.969
stomach	0.876	0.797	0.946	0.944	0.938	0.946
aorta	0.929	0.909	0.929	0.931	0.925	0.932
inferior vena cava	0.834	0.827	0.854	0.842	0.729	0.856
portal vein and splenic vein	0.649	0.752	0.781	0.775	0.734	0.780
pancreas	0.759	0.820	0.807	0.841	0.797	0.853
right adrenal gland	0.604	0.661	0.696	0.692	0.673	0.699
left adrenal gland	0.638	0.642	0.643	0.646	0.666	0.660
BTCV-Cervix						
bladder	0.730	0.752	0.785	0.800	0.863	0.871
prostate or uterus	0.714	0.675	-	0.587	0.691	0.714
rectum	0.719	0.688	-	-	-	-
small bowel	0.466	0.527	0.437	0.544	0.608	0.679
VerSe						
vertebrae C1	0.795	0.862	0.875	0.859	0.844	0.863
vertebrae C2	0.867	0.852	0.909	0.881	0.862	0.890
vertebrae C3	0.804	0.844	0.882	0.828	0.863	0.869
vertebrae C4	0.796	0.874	0.877	0.857	0.811	0.868
vertebrae C5	0.794	0.855	0.878	0.851	0.861	0.864
vertebrae C6	0.808	0.816	0.877	0.865	0.863	0.874
vertebrae C7	0.798	0.822	0.892	0.857	0.878	0.887
vertebrae T1	0.832	0.800	0.901	0.847	0.898	0.897
vertebrae T2	0.817	0.840	0.887	0.862	0.899	0.905
vertebrae T3	0.808	0.837	0.836	0.848	0.892	0.894
vertebrae T4	0.777	0.775	0.790	0.844	0.896	0.903
vertebrae T5	0.745	0.794	0.776	0.827	0.898	0.908
vertebrae T6	0.713	0.782	0.766	0.818	0.905	0.913
vertebrae T7	0.723	0.887	0.742	0.822	0.912	0.919
vertebrae T8	0.710	0.847	0.759	0.791	0.912	0.920
vertebrae T9	0.722	0.826	0.810	0.804	0.916	0.925
vertebrae T10	0.770	0.852	0.803	0.786	0.922	0.928
vertebrae T11	0.776	0.837	0.820	0.822	0.926	0.932
vertebrae T12	0.835	0.798	0.879	0.870	0.927	0.931
vertebrae L1	0.873	0.871	0.915	0.864	0.930	0.936
vertebrae L2	0.822	0.800	0.871	0.811	0.929	0.932
vertebrae L3	0.787	0.876	0.798	0.752	0.927	0.928

vertebrae L4	0.755	0.773	0.722	0.707	0.930	0.932
vertebrae L5	0.740	0.763	0.716	0.735	0.913	0.919
vertebrae L6	0.434	0.475	-	-	-	-
AbdomenCT-1K						
liver	0.978	0.982	0.969	0.974	0.896	0.976
kidney	0.947	0.944	0.912	-	-	-
spleen	0.967	0.976	0.968	0.966	0.959	0.964
pancreas	0.857	0.860	0.828	0.865	0.853	0.881
AMOS22						
spleen	0.953	0.946	0.930	0.934	0.933	0.946
right kidney	0.955	0.943	0.940	0.945	0.937	0.949
left kidney	0.944	0.950	0.925	0.931	0.938	0.948
gallbladder	0.779	0.832	0.813	0.847	0.814	0.855
esophagus	0.805	0.808	0.777	0.776	0.783	0.805
liver	0.971	0.972	0.958	0.959	0.901	0.960
stomach	0.858	0.855	0.882	0.876	0.863	0.889
aorta	0.944	0.953	0.914	0.917	0.897	0.921
inferior vena cava	0.889	0.870	0.809	0.855	0.669	0.865
pancreas	0.809	0.840	0.773	0.797	0.757	0.828
right adrenal gland	0.744	0.708	0.683	0.700	0.657	0.721
left adrenal gland	0.740	0.714	0.684	0.704	0.687	0.724
duodenum	0.743	0.754	0.639	0.704	0.337	0.729
bladder	0.824	0.808	0.809	0.826	0.819	0.847
prostate or uterus	0.817	0.827	-	0.788	0.790	0.828
TotalSegmentatorV2						
spleen	0.957	0.969	0.982	0.967	0.965	0.971
right kidney	0.949	0.940	0.962	0.934	0.930	0.948
left kidney	0.942	0.922	0.961	0.920	0.921	0.941
gallbladder	0.807	0.843	0.896	0.827	0.782	0.833
liver	0.964	0.965	0.982	0.968	0.944	0.974
stomach	0.929	0.935	0.960	0.931	0.917	0.939
aorta	0.954	0.961	0.961	0.959	0.949	0.965
inferior vena cava	0.892	0.902	0.896	0.883	0.695	0.896
portal vein and splenic vein	0.757	0.830	0.835	0.801	0.744	0.818
pancreas	0.845	0.856	0.917	0.860	0.833	0.877
right adrenal gland	0.805	0.877	0.909	0.863	0.834	0.869
left adrenal gland	0.808	0.866	0.914	0.873	0.851	0.881
left lung upper lobe	0.943	0.939	0.979	0.953	0.931	0.955
left lung lower lobe	0.928	0.953	0.964	0.938	0.899	0.944
right lung upper lobe	0.896	0.912	0.919	0.878	0.872	0.905
right lung middle lobe	0.905	0.939	0.952	0.916	0.909	0.930
right lung lower lobe	0.928	0.950	0.974	0.943	0.893	0.951
vertebrae L5	0.909	0.930	0.946	0.916	0.916	0.933
vertebrae L4	0.899	0.929	0.947	0.899	0.917	0.933
vertebrae L3	0.892	0.927	0.967	0.925	0.934	0.957
vertebrae L2	0.925	0.928	0.975	0.936	0.950	0.968
vertebrae L1	0.904	0.917	0.967	0.919	0.934	0.955
vertebrae T12	0.902	0.912	0.961	0.902	0.930	0.948
vertebrae T11	0.899	0.922	0.970	0.900	0.930	0.952
vertebrae T10	0.900	0.918	0.972	0.901	0.937	0.955
vertebrae T9	0.886	0.918	0.976	0.901	0.936	0.960
vertebrae T8	0.882	0.893	0.967	0.872	0.913	0.949
vertebrae T7	0.822	0.886	0.920	0.831	0.890	0.920
vertebrae T6	0.840	0.902	0.943	0.878	0.910	0.933
vertebrae T5	0.869	0.923	0.944	0.891	0.904	0.930
vertebrae T4	0.876	0.910	0.948	0.887	0.910	0.935
vertebrae T3	0.888	0.926	0.950	0.895	0.903	0.935
vertebrae T2	0.909	0.918	0.967	0.920	0.922	0.949
vertebrae T1	0.907	0.945	0.969	0.933	0.926	0.950
vertebrae C7	0.894	0.943	0.964	0.923	0.901	0.937
vertebrae C6	0.839	0.840	0.941	0.882	0.864	0.917
vertebrae C5	0.797	0.852	0.915	0.825	0.852	0.862
vertebrae C4	0.860	0.859	0.944	0.904	0.881	0.917
vertebrae C3	0.857	0.936	0.956	0.905	0.905	0.926
vertebrae C2	0.908	0.953	0.972	0.910	0.872	0.933
vertebrae C1	0.884	0.862	0.935	0.894	0.848	0.896
esophagus	0.874	0.913	0.952	0.907	0.886	0.916

trachea	0.926	0.945	0.974	0.941	0.910	0.946
brain	0.870	0.946	0.943	0.894	0.892	0.903
left iliac artery	0.822	0.896	0.916	0.895	0.872	0.906
right iliac artery	0.820	0.879	0.915	0.875	0.877	0.899
left iliac vena	0.841	0.898	0.941	0.917	0.899	0.925
right iliac vena	0.834	0.884	0.919	0.890	0.846	0.908
small bowel	0.854	0.868	0.918	0.834	0.840	0.865
duodenum	0.779	0.805	0.900	0.822	0.596	0.848
colon	0.882	0.882	0.948	0.898	0.819	0.906
left rib 1	0.914	0.938	0.948	0.909	0.875	0.918
left rib 2	0.934	0.927	0.966	0.932	0.909	0.943
left rib 3	0.906	0.929	0.950	0.910	0.885	0.907
left rib 4	0.908	0.936	0.947	0.903	0.887	0.927
left rib 5	0.878	0.895	0.933	0.889	0.889	0.928
left rib 6	0.865	0.912	0.925	0.866	0.884	0.916
left rib 7	0.885	0.907	0.942	0.877	0.901	0.934
left rib 8	0.902	0.888	0.955	0.890	0.910	0.941
left rib 9	0.910	0.901	0.953	0.897	0.916	0.944
left rib 10	0.911	0.883	0.949	0.893	0.906	0.937
left rib 11	0.891	0.894	0.949	0.903	0.911	0.938
left rib 12	0.885	0.873	0.912	0.883	0.871	0.909
right rib 1	0.905	0.938	0.945	0.907	0.875	0.912
right rib 2	0.933	0.946	0.959	0.924	0.888	0.929
right rib 3	0.906	0.938	0.931	0.891	0.854	0.900
right rib 4	0.928	0.942	0.949	0.906	0.882	0.926
right rib 5	0.905	0.893	0.916	0.876	0.877	0.914
right rib 6	0.900	0.929	0.951	0.886	0.907	0.932
right rib 7	0.903	0.914	0.960	0.884	0.915	0.942
right rib 8	0.888	0.928	0.959	0.887	0.913	0.941
right rib 9	0.892	0.928	0.950	0.890	0.920	0.946
right rib 10	0.900	0.927	0.949	0.896	0.916	0.945
right rib 11	0.880	0.924	0.933	0.885	0.891	0.924
right rib 12	0.885	0.906	0.917	0.883	0.880	0.907
left humerus	0.911	0.867	0.930	0.854	0.881	0.903
right humerus	0.916	0.794	0.940	0.873	0.884	0.913
left scapula	0.910	0.949	0.959	0.911	0.887	0.921
right scapula	0.916	0.923	0.959	0.922	0.887	0.920
left clavicle	0.955	0.917	0.975	0.952	0.931	0.956
right clavicle	0.937	0.940	0.973	0.945	0.933	0.952
left femur	0.944	0.882	0.970	0.940	0.944	0.954
right femur	0.944	0.911	0.980	0.945	0.957	0.959
left hip	0.944	0.937	0.975	0.947	0.938	0.955
right hip	0.939	0.932	0.986	0.950	0.961	0.959
sacrum	0.925	0.933	0.958	0.915	0.895	0.922
left gluteus maximus	0.925	0.927	0.977	0.940	0.938	0.949
right gluteus maximus	0.917	0.930	0.978	0.937	0.937	0.949
left gluteus medius	0.919	0.926	0.973	0.931	0.923	0.923
right gluteus medius	0.908	0.927	0.978	0.938	0.937	0.946
left gluteus minimus	0.875	0.917	0.965	0.914	0.903	0.919
right gluteus minimus	0.876	0.920	0.967	0.915	0.896	0.921
left autochthon	0.939	0.934	0.978	0.951	0.932	0.953
right autochthon	0.941	0.932	0.976	0.941	0.927	0.947
left iliopsoas	0.876	0.910	0.965	0.921	0.898	0.926
right iliopsoas	0.876	0.916	0.952	0.907	0.898	0.914
bladder	0.890	0.906	0.934	0.899	0.895	0.915
left atrial appendage	0.863	0.900	0.942	0.901	0.873	0.910
brachiocephalic trunk	0.872	0.899	0.936	0.892	0.888	0.915
left brachiocephalic vein	0.881	0.919	0.945	0.903	0.885	0.898
right brachiocephalic vein	0.862	0.909	0.922	0.884	0.869	0.901
left common carotid artery	0.826	0.884	0.925	0.868	0.828	0.891
right common carotid artery	0.755	0.858	0.885	0.811	0.784	0.844
costal cartilages	0.844	0.868	0.888	0.856	0.833	0.864
heart	0.932	0.928	0.937	0.919	0.916	0.924
left kidney cyst	0.623	0.858	0.892	0.618	0.752	0.858
right kidney cyst	0.568	0.841	0.716	0.606	0.615	0.681
prostate	0.743	0.752	0.808	0.744	0.745	0.774
pulmonary vein	0.838	0.820	0.916	0.830	0.847	0.863

skull	0.909	0.849	0.893	0.827	0.769	0.857
spinal cord	0.911	0.950	0.959	0.934	0.905	0.937
sternum	0.896	0.906	0.897	0.899	0.884	0.911
left subclavian artery	0.833	0.901	0.929	0.877	0.857	0.892
right subclavian artery	0.818	0.870	0.916	0.861	0.850	0.885
superior vena cava	0.894	0.899	0.932	0.888	0.905	0.923
thyroid gland	0.832	0.886	0.908	0.866	0.853	0.890
vertebrae S1	0.870	0.906	0.925	0.890	0.880	0.909

References

- [1] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. The medical segmentation decathlon. *Nature communications*, 13(1):4128, 2022. 1
- [2] SG Armato III, G McLennan, L Bidaut, MF McNitt-Gray, CR Meyer, AP Reeves, B Zhao, DR Aberle, CI Henschke, EA Hoffman, et al. Data from LIDC-IDRI [data set]. The Cancer Imaging Archive, 2015. 1
- [3] Yuxin Du, Fan Bai, Tiejun Huang, and Bo Zhao. Segvol: Universal and interactive volumetric medical image segmentation. *arXiv preprint arXiv:2311.13385*, 2023. 3
- [4] Wasserthal et al. TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023. 1, 2
- [5] Stephanie A Harmon, Thomas H Sanford, Sheng Xu, Evrim B Turkbey, Holger Roth, Ziyue Xu, Dong Yang, Andriy Myronenko, Victoria Anderson, Amel Amalou, et al. Artificial intelligence for the detection of covid-19 pneumonia on chest ct using multinational datasets. *Nature communications*, 11(1):4080, 2020. 1
- [6] Yufan He, Pengfei Guo, Yucheng Tang, Andriy Myronenko, Vishwesh Nath, Ziyue Xu, Dong Yang, Can Zhao, Daguang Xu, and Wenqi Li. A short review and evaluation of sam2’s performance in 3d ct image segmentation. *arXiv preprint arXiv:2408.11210*, 2024. 3
- [7] Yuanfeng Ji, Haotian Bai, Jie Yang, Chongjian Ge, Ye Zhu, Ruimao Zhang, Zhen Li, Lingyan Zhang, Wanling Ma, Xiang Wan, et al. AMOS: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. *arXiv preprint arXiv:2206.08023*, 2022. 1
- [8] C Daniel Johnson, Mei-Hsiu Chen, Alicia Y Toledano, Jay P Heiken, Abraham Dachman, Mark D Kuo, Christine O Menias, Betina Siewert, Jugesh I Cheema, Richard G Obregon, et al. Accuracy of CT colonography for detection of large adenomas and cancers. *New England Journal of Medicine*, 359(12):1207–1217, 2008. 1
- [9] Jun Ma, Yao Zhang, Song Gu, Cheng Zhu, Cheng Ge, Yichi Zhang, Xingle An, Congcong Wang, Qiyuan Wang, Xin Liu, et al. Abdomenct-1k: Is abdominal organ segmentation a solved problem? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6695–6714, 2021. 1
- [10] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15:1–9, 2024. 3
- [11] Blaine Rister, Darvin Yi, Kaushik Shivakumar, Tomomi Nobashi, and Daniel L Rubin. CT-ORG, a new dataset for multiple organ segmentation in computed tomography. *Scientific Data*, 7(1):381, 2020. 1
- [12] H Roth, A Farag, EB Turkbey, L Lu, J Liu, and RM Summers. Data from pancreas-CT (version 2)[data set]. The Cancer Imaging Archive (2016), 2016. 1
- [13] Joel Saltz, Mary Saltz, Prateek Prasanna, Richard Moffitt, Janos Hajagos, Erich Bremer, Joseph Balsamo, and Tahsin Kurc. Stony Brook university COVID-19 positive cases. *the cancer imaging archive*, 4, 2021. 1
- [14] Anjany Sekuboyina, Malek E Hussein, Amirhossein Bayat, Maximilian Löffler, Hans Liebl, Hongwei Li, Giles Tetteh, Jan Kukačka, Christian Payer, Darko Štern, et al. Verse: A vertebrae labelling and segmentation benchmark for multi-detector CT images. *Medical image analysis*, 73:102166, 2021. 1
- [15] Amber L Simpson, Jacob Peoples, John M Creasy, Gabor Fichtinger, Natalie Gangai, Krishna N Keshavamurthy, Andras Lasso, Jinru Shia, Michael I D’Angelica, and Richard KG Do. Preoperative CT and survival data for patients undergoing resection of colorectal liver metastases. *Scientific Data*, 11(1):172, 2024. 1
- [16] Karen-Helene Støverud, David Bouget, Andre Pedersen, Håkon Olav Leira, Thomas Langø, and Erlend Fager-tun Hofstad. AeroPath: An airway segmentation benchmark dataset with challenging pathology. *arXiv preprint arXiv:2311.01138*, 2023. 1
- [17] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5):395–409, 2011. 1