

Supplementary

Related work

The Segment Anything Model (SAM) [1] is the first foundation model in image segmentation. However, the applicability of SAM to medical image segmentation remains limited due to the significant differences between natural images and medical images. SAM is a promptable segmentation method that requires points or bounding boxes to specify the segmentation targets. This resembles conventional interactive segmentation methods [2]–[5] but SAM has better generalization ability, while existing deep learning-based interactive segmentation methods focus mainly on limited tasks and image modalities. Several studies have shown that SAM can fail in typical medical image segmentation tasks [6]–[13] and other challenging scenarios [14]–[17], especially when targets have weak boundaries. This motivates us to develop a universal segmentation tool that can segment a wide range of anatomical structures and medical imaging modalities. While concurrent studies [18]–[20] also tried to adapt SAM for the medical domain, our work stands out as the first pre-trained model specifically for general medical image segmentation tasks.

TABLE 1

Computed Tomography (CT) image datasets. Datasets marked with * denote external validation sets and the remaining datasets are used for internal validation.

Dataset Name	Modality	Segmentation Targets	# of scans
AbdomenCT-1K [21], [22]	CT	Liver, kidneys, pancreas, spleen	1056
AMOS-CT [23]	CT	Abdominal organ	200
AutoPET [24]	PET-CT	Whole-body tumor	900
COVID-19 Seg. Challenge [25], [26]	CT	COVID-19 infections	199
COVID-19-CT-Seg [27]	CT	COVID-19 infections, left lung, and right lung	20
GLIS-RT [28]	CT	Head tumor	75
HCC-TACE-Seg* [29], [30]	CT	Liver cancer	70
HECKTOR [31]	PET-CT	Head and neck tumor	524
INSTANCE [32]	CT	Hematoma	100
KiPA [33], [34]	CT	Kidney, tumor, renal artery, renal vein	70
KITS [35]	CT	Kidney, tumor, cyst	489
Lymph Nodes [36], [37]	CT	Lymph nodes	176
MMWHS-CT [38]–[41]	CT	Heart anatomies	20
MSD-Colon Tumor [42]	CT	Colon tumor	126
MSD-Hepatic Tumor [42]	CT	Hepatic tumor	303
MSD-Lung Tumor [42]	CT	Lung tumor	96
MSD-Pancreas [42]	CT	Pancreas, pancreas tumor	281
MSD-Spleen [42]	CT	Spleen	61
NSCLC Pleural Effusion [29], [43], [44]	CT	Pleural effusion	78
NSCLC Radiogenomics [45]	CT	Lung Tumor	88
ORG [46]	CT	Whole-body organs	140
SegTHOR [47]	CT	Esophagus, heart, aorta, trachea	40
StructSeg* [48]	CT	Nasopharyngeal cancer and lung cancer, with OAR and GTV	50
TotalSegmentator [49]	CT	Whole body organs	1204
WORD* [50]	CT	Abdominal organs	150

TABLE 2

Magnetic Resonance (MR) image datasets. Datasets marked with * denote external validation sets and the remaining datasets are used for internal validation.

Dataset Name	Modality	Segmentation Targets	# of scans
ACDC* [51]	MR	Heart anatomies	150
AMOS-MR [23]	MR	Abdominal organ	40
ATLAS R2.0 [52]	MR-T1	Brain stroke	1271
Brain Tumor Dataset Figshare [53], [54]	MR-T1ce	Brain tumor	233
Brain TR-GammakKnife [55]	MR	Brain lesion	47
BraTS [56]–[60]	MR-T1, MR-T1CE, MR-T2, MR-FLAIR	Brain tumor	1251
CC-Tumor Heterogeneity* [61]	MR	Cervical cancer	7
CHAOS* [62]	MR-T1, MR-T2	Liver, kidney, spleen	60
crossMoDA [63]	MR	Brain tumor	227
FeTA [63]	MR-Fetal	Brain tissues	160
HaN-Seg* [64]	MR	Head organs	42
ISLES [65]	MR-DWI, MR-ADC, MR-FLAIR	Ischemic stroke lesion	180
I2CVB [66]	MR-T2, MR-DWI	Prostate	19
Meningioma-SEG-CLASS [67]	MR-T1ce, T2-FLAIR	Tumor (meningioma)	191
MMs [68]	MR	Heart anatomies	150
MMWHS-MR [38]–[41]	MR	Heart anatomies	20
MSD-Heart [69]	MR	Left atrial	30
MSD-Prostate [42]	MR-ADC, MR-T2	Prostate	48
NCI-ISBI [70]	MR-ADC, MR-T2	Prostate	48
PI-CAI [71]	MR-bp	Prostate cancer	1584
PPMI [72]	MR-T1	Brain regions of Parkinson patients	1130
PROMISE [73]	MR-T2	Prostate	50
Qin-Prostate-Repeatability [42], [74]	MR	Left atrium	30
QUBIQ* [75]	MR	Prostate	52
Spine [76]	MR	Vertebrae	172
WMH [77]	MR-T1, MR-FLAIR	White matter hyper-intensities	60

TABLE 3

Chest X-ray (CXR), Mammography, Optical Coherence Tomography (OCT), and Ultrasound image datasets. Datasets marked with * denote external validation sets and the remaining datasets are used for internal validation.

Dataset Name	Modality	Segmentation Targets	# of images
Chest Xray Masks and Labels [78] [79]	CXR	Lung	704
Chest X-Ray (Pneumothorax) [80] [81]	CXR	Pneumothorax	2668
COVID-19 Radiography* [82] [83]	CXR	Lung, COVID-19 infection, lung opacity, viral pneumonia	21165
COVID-QU-Ex [84] [82] [85] [83] [86]	CXR	Lung, COVID infection	5826
JSRT [87]	CXR	Lung, heart	307
Lung [88] [89] [90]	CXR	Lung	30330
QaTa-COV19 [91]	CXR	COVID-19 infection	9258
CDD-CESM [29] [92] [93]	Mammography	Abnormal findings in breast tissue	1233
Intraretinal Cystoid Fluid [94] [95]	OCT	Cystoid macular edema	1460
OCT Images (DME) [96]	OCT	Diabetic macular edema	610
AbdomenUS [97]	Ultrasound	Gallbladder, kidney, liver, spleen, vessel	60
Breast Cancer [98]	Ultrasound	Benign & malignant breast lesion	647
CAMUS [99]	Ultrasound	Left ventricle endocardium, left atrium	21232
CT2USforKidneySeg [100]	Ultrasound	Kidney	4586
FH-PS-AOP [101]	Ultrasound	Pubic symphysis, fetal head	4000
HC [102]	Ultrasound	Fetal head circumference	999
TN-SCUI [103]	Ultrasound	Benign & malignant thyroid nodules	3644
Nerve [104]	Ultrasound	Brachial plexus	5635

TABLE 4

Dermoscopy, Endoscopy, Fundus, and Pathology image datasets. Datasets marked with * denote external validation sets and the remaining datasets are used for internal validation.

Dataset Name	Modality	Segmentation Targets	# of images
ISIC [105] [106] [107]	Dermoscopy	Skin cancer	3694
UWaterloo Skin Cancer* [108]	Dermoscopy	Skin cancer	206
BKAI-IGH NeoPolyp [109]	Endoscopy	Polyp	1000
CholecSeg8k [110] [111]	Endoscopy	Surgical tools and abdominal tissues	8080
Kvasir* [112]	Endoscopy	Polyp	1000
m2caiSeg [113]	Endoscopy	Surgical tools and abdominal tissues	307
PolypGen [114] [115] [116]	Endoscopy	Polyp	1412
RoboTool [117]	Endoscopy	Surgical tools	500
sisvse [118]	Endoscopy	Surgical tools and abdominal tissues	18218
IDRiD* [119]	Fundus	Optic disc	81
PAPILA* [120]	Fundus	Optic disc and cup	488
REFUGE [121] [122]	Fundus	Optic disc and cup	1200
GlaS@MICCAI2015* [123] [124]	Pathology	Adenocarcinomas	165
HuBMAP HPA [125]	Pathology	Function tissue unit (FTU) of large intestine	58
HuBMAP Hacking the Kidney [126]	Pathology	Glomeruli FTU	8

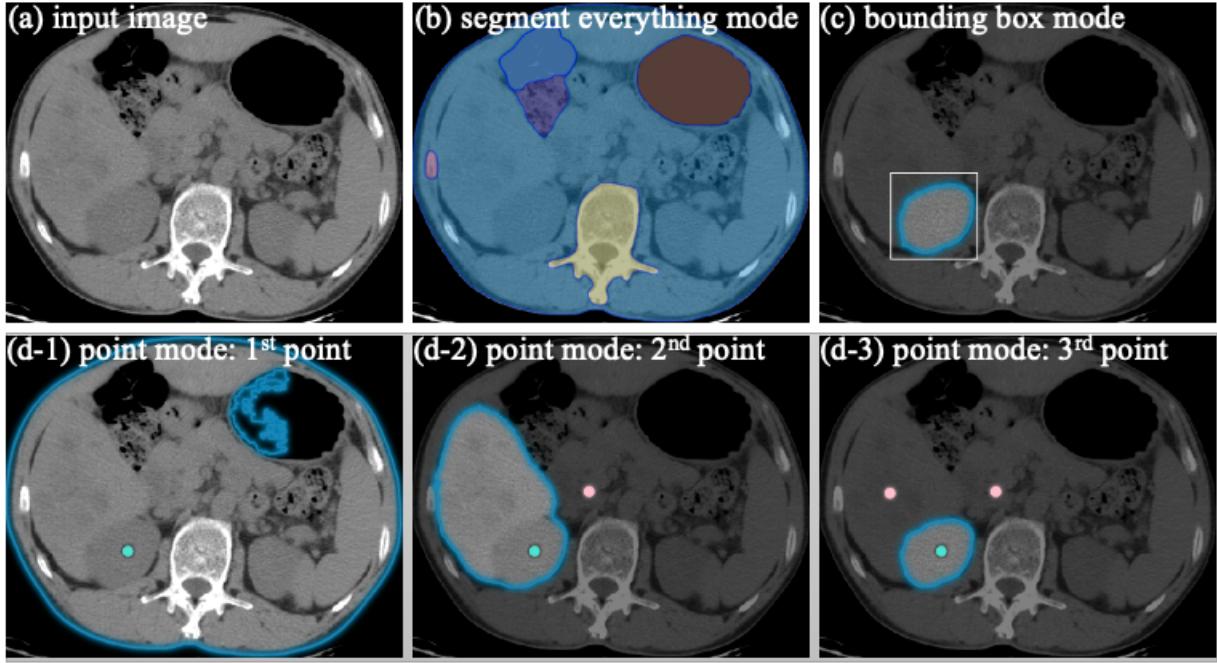


Fig. 1. Segmentation results of SAM based on different segmentation modes. The segment-everything mode divides the whole image into six regions based on the image intensity (b). However, such segmentation results have limited use because of two main reasons. On the one hand, the segmentation results do not have semantic labels. On the other hand, clinicians mainly focus on meaningful ROIs in clinical scenarios, e.g., the liver, kidneys, spleen, and lesions. The bounding box-based segmentation mode generates good results for the right kidney by just giving the upper-left and bottom-right points (c). For the point-based segmentation mode (d), we first give one foreground point to the center of the right kidney but the segmentation results include the whole abdomen tissues. Then, we add a background point on the over-segmentation regions. The segmentation mask shrinks to the liver and right kidney. After adding another background point on the liver, we finally obtain the expected kidney segmentation.

TABLE 5

Internal validation results of SAM, specialist U-Net, and MedSAM on CT image segmentation tasks in terms of DSC and NSD. Scores are displayed as Median values (First quartile, Third quartile). Scores marked with * indicate statistically significant superiority over SAM, while those marked with † denote statistically significant superiority of MedSAM over the specialist U-Net model (p -value < 0.05).

Target	Modality	DSC (%)			NSD (%)		
		SAM	U-Net	MedSAM	SAM	U-Net	MedSAM
Aorta	CT	91.9 (89.3, 93.4)	93.1 (91.2, 94.9)*	95.6 (95.1, 96.1)*†	93.6 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.4)*†
COVID-19 Infection	CT	62.4 (45.5, 72.5)	71.3 (49.8, 81.7)*	84.2 (76.4, 90.1)*†	65.6 (48.8, 86.7)	67.2 (48.4, 89.5)	93.2 (74.2, 97.9)*†
Colon Cancer	CT	72.6 (64.5, 81.3)	80.3 (75.7, 83.9)*	83.0 (77.7, 86.4)*†	87.2 (86.5, 93.1)	93.7 (92.6, 95.2)*	97.4 (96.7, 98.1)*†
Esophagus	CT	71.5 (68.1, 75.5)	56.3 (51.4, 63.6)*	73.7 (68.8, 78.4)*†	91.4 (83.4, 94.4)	84.2 (55.6, 94.2)*	95.6 (86.8, 97.9)*†
Gallbladder	CT	85.4 (77.4, 90.1)	87.0 (82.8, 92.1)	91.8 (87.9, 95.2)*†	93.5 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.4)*†
Glioblastoma	CT	74.4 (69.8, 78.2)	86.2 (83.4, 88.4)*	94.3 (93.3, 95.4)*†	83.3 (81.9, 85.3)	89.4 (86.2, 91.5)*	99.4 (99.2, 99.6)*†
Head-Neck Cancer	CT	61.4 (44.2, 72.1)	67.3 (58.5, 75.1)*	79.4 (71.9, 83.8)*†	75.5 (69.1, 81.7)	70.7 (62.5, 80.1)*	90.5 (85.7, 94.8)*†
Heart	CT	91.5 (89.8, 93.8)	94.2 (93.4, 95.6)*	97.6 (97.4, 97.9)*†	94.0 (92.4, 95.0)	85.6 (83.2, 88.2)*	96.3 (95.3, 97.3)*†
Inferior Vena Cava	CT	82.9 (78.6, 86.1)	85.4 (82.1, 88.6)*	91.8 (88.9, 93.6)*†	93.6 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.4)*†
Intracranial Hemorrhage	CT	86.7 (82.9, 91.6)	89.0 (83.4, 93.4)*	94.0 (91.5, 94.9)*†	94.9 (91.0, 98.6)	98.6 (96.6, 99.6)*	99.4 (99.1, 99.7)*†
Kidney Cancer	CT	86.7 (82.6, 90.3)	87.7 (80.9, 92.0)	90.2 (86.5, 93.7)*†	93.1 (90.9, 95.8)	93.3 (87.6, 96.6)	98.7 (97.5, 99.4)*†
Left Adrenal Gland	CT	60.0 (52.0, 69.0)	52.2 (49.6, 58.3)*	65.4 (59.9, 74.4)†	93.6 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.4)*†
Left Kidney	CT	94.6 (92.5, 96.6)	95.5 (94.4, 96.6)	97.0 (96.6, 97.7)*†	93.5 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.3)*†
Left Lung	CT	94.4 (92.3, 95.6)	95.6 (93.8, 96.5)*	95.7 (95.0, 96.3)*†	94.2 (90.4, 96.6)	94.9 (92.2, 96.8)*	98.0 (96.4, 98.9)*†
Liver	CT	93.2 (91.6, 94.7)	96.6 (96.1, 97.1)*	98.0 (97.8, 98.4)*†	93.6 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.4)*†
Liver Cancer	CT	80.0 (72.6, 83.8)	85.2 (80.1, 90.9)*	88.7 (84.2, 93.0)*†	90.1 (86.3, 93.3)	93.3 (89.6, 96.6)*	97.9 (96.2, 98.5)*†
Lung Cancer	CT	73.2 (64.2, 81.6)	85.5 (78.9, 88.3)*	94.1 (90.4, 95.5)*†	89.4 (84.4, 93.1)	91.9 (88.8, 94.7)*	99.5 (98.7, 99.7)*†
Lymph Nodes	CT	77.1 (74.1, 82.1)	69.2 (67.8, 77.3)*	82.1 (78.8, 87.6)*†	90.2 (85.7, 93.0)	79.6 (75.3, 84.1)*	93.4 (91.9, 96.4)*†
Melanoma	CT	63.7 (54.5, 70.6)	78.0 (73.3, 82.9)*	83.8 (79.6, 87.6)*†	80.4 (77.3, 83.7)	86.6 (83.9, 91.1)*	94.0 (92.3, 95.4)*†
Pancreas	CT	73.1 (66.9, 76.9)	80.3 (76.1, 82.1)*	87.2 (85.9, 90.5)*†	93.6 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.4)*†
Pancreas Cancer	CT	74.1 (69.1, 77.2)	69.9 (62.1, 80.3)*	79.1 (75.5, 83.2)*†	84.7 (81.1, 88.6)	80.5 (76.2, 89.2)*	95.0 (92.1, 97.5)*†
Pleural Effusion	CT	35.3 (21.7, 53.7)	70.9 (62.5, 80.4)*	82.1 (71.6, 92.9)*†	61.3 (60.6, 71.5)	80.2 (77.9, 86.8)*	95.2 (86.9, 98.6)*†
Right Adrenal Gland	CT	51.5 (47.9, 59.6)	49.8 (44.9, 57.5)	66.8 (57.7, 72.6)*†	93.6 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.4)*†
Right Kidney	CT	94.8 (92.3, 96.6)	95.8 (94.3, 96.8)*	97.1 (96.4, 97.9)*†	93.6 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.4)*†
Right Lung	CT	95.2 (92.6, 96.0)	95.8 (94.1, 96.7)*	97.1 (96.6, 97.7)*†	94.2 (90.4, 96.6)	94.9 (92.2, 96.8)	98.0 (96.4, 98.9)*†
Spinal Cord	CT	75.4 (70.9, 78.9)	54.2 (51.4, 57.9)*	78.1 (75.1, 80.2)*†	87.5 (81.6, 94.0)	59.2 (53.8, 85.0)*	90.0 (84.7, 95.9)*†
Spleen	CT	94.3 (92.3, 96.5)	96.0 (95.1, 97.4)*	97.6 (97.2, 98.2)*†	93.5 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.3)*†
Stomach	CT	86.3 (78.3, 90.2)	92.8 (90.7, 94.3)*	96.2 (95.1, 97.2)*†	93.6 (91.8, 94.8)	97.2 (96.1, 97.9)*	99.0 (98.5, 99.4)*†
Throat Cancer	CT	28.1 (20.9, 53.6)	66.6 (66.4, 78.1)*	80.3 (79.3, 86.9)*†	74.4 (69.1, 92.1)	85.5 (80.7, 90.4)	98.1 (98.1, 99.5)*†

TABLE 6

Internal validation results of SAM, specialist U-Net, and MedSAM on MR image segmentation tasks in terms of DSC and NSD. Scores are displayed as Median values (First quartile, Third quartile). Scores marked with * indicate statistically significant superiority over SAM, while those marked with † denote statistically significant superiority of MedSAM over the specialist U-Net model (p -value < 0.05).

Target	Modality	DSC (%)			NSD (%)		
		SAM	U-Net	MedSAM	SAM	U-Net	MedSAM
Brainstem	MR	69.2 (65.4, 75.8)	89.9 (88.7, 91.1)*	97.1 (96.7, 97.2)*†	62.6 (57.3, 67.9)	93.8 (92.9, 97.5)*	100 (99.6, 100)*†
Cerebellum	MR	76.5 (72.1, 78.7)	91.0 (86.9, 91.5)*	96.8 (96.3, 97.2)*†	58.7 (52.1, 64.5)	91.8 (86.1, 93.5)*	99.8 (99.1, 99.9)*†
Deep Grey Matter	MR	49.6 (25.4, 60.0)	80.1 (74.4, 87.8)*	95.6 (92.7, 95.9)*†	50.4 (48.3, 54.0)	76.9 (69.1, 80.8)*	96.6 (94.4, 98.6)*†
Glioma	MR-FLAIR	83.4 (78.9, 87.0)	92.8 (90.2, 94.5)*	96.2 (94.4, 97.1)*†	70.5 (62.4, 78.6)	93.9 (90.7, 96.9)*	98.8 (97.0, 99.7)*†
Glioma	MR-T1	76.3 (70.6, 80.3)	87.8 (84.2, 90.5)*	94.4 (91.6, 95.8)*†	56.1 (50.7, 62.8)	82.3 (76.2, 87.7)*	97.6 (93.9, 99.0)*†
Glioma Enhancing Tumor	MR-T2	78.8 (73.8, 83.8)	86.9 (82.2, 90.4)*	95.2 (92.7, 96.4)*†	68.8 (60.8, 76.7)	84.7 (79.1, 88.9)*	99.0 (96.2, 99.8)*†
Glioma Tumor Core	MR-T1-CE	71.0 (62.5, 79.1)	90.5 (84.7, 94.2)*	95.9 (92.7, 97.3)*†	74.4 (67.1, 83.8)	97.9 (92.8, 99.5)*	99.8 (98.9, 100)*†
Ischemic Stroke	MR-ADC	61.3 (54.4, 68.9)	82.9 (75.2, 87.5)*	92.3 (87.3, 95.3)*†	75.9 (69.9, 81.0)	95.6 (90.4, 98.7)*	94.5 (91.2, 98.0)*
Left Atrium	MR	83.6 (82.8, 83.7)	90.8 (90.3, 91.3)	97.3 (97.1, 97.6)	70.2 (67.3, 73.2)	90.6 (87.8, 92.7)	99.1 (98.4, 99.4)
Left Ventricle	MR	77.5 (76.1, 81.8)	94.5 (93.6, 95.7)*	98.5 (97.9, 98.7)*†	85.1 (79.2, 88.8)	99.6 (99.3, 99.8)*	100 (100, 100)*†
Meningioma	MR-T1-CE	92.1 (87.4, 93.5)	94.6 (90.3, 95.2)*	97.9 (97.6, 98.1)*†	95.5 (92.6, 97.6)	98.6 (96.7, 99.7)*	100 (100, 100)*†
Meningioma	MR-T2-FLAIR	79.2 (72.8, 85.4)	89.2 (88.7, 92.1)*	97.0 (96.2, 97.6)*†	66.4 (61.3, 83.5)	94.7 (90.6, 98.6)*	100 (99.4, 100)*†
Prostate	MR-ADC	83.7 (81.0, 85.1)	91.1 (89.7, 92.6)*	98.5 (98.2, 98.5)*†	83.2 (82.0, 85.4)	96.0 (93.5, 97.4)*	100 (100, 100)*†
Prostate Cancer	MR	69.3 (64.4, 72.2)	79.2 (74.8, 81.8)*	96.9 (93.1, 97.5)*†	78.1 (73.6, 81.2)	89.8 (84.6, 94.0)*	100 (99.3, 100)*†
Prostate	MR-T2	90.7 (89.0, 91.5)	93.8 (93.0, 94.4)*	98.4 (98.3, 98.6)*†	90.1 (88.0, 92.7)	97.2 (95.3, 98.0)*	100 (100, 100)*†
Right Ventricle	MR	90.3 (88.1, 92.5)	93.9 (91.7, 95.2)*	97.2 (96.7, 98.1)*†	95.6 (91.1, 98.6)	99.7 (99.3, 100)*	100 (100, 100)*†
Spine	MR	80.8 (77.4, 84.0)	87.7 (85.6, 88.6)*	91.8 (90.9, 93.1)*†	86.0 (83.5, 89.2)	92.9 (90.9, 93.7)*	96.8 (96.1, 97.6)*†
Ventricles	MR	63.9 (47.9, 69.8)	69.5 (56.0, 73.4)*	90.0 (79.4, 91.9)*†	63.0 (56.2, 66.4)	66.8 (61.8, 68.8)*	85.1 (78.7, 90.7)*†
Vestibular Schwannoma	MR	85.3 (78.5, 89.5)	90.3 (80.3, 93.2)*	95.2 (91.6, 97.7)*†	97.0 (94.5, 97.9)	99.2 (96.7, 99.9)*	99.9 (98.4, 100)*†

TABLE 7

Internal validation results of SAM, specialist U-Net, and MedSAM on grey images (Chest X-Ray (CXR), Mammography, OCT, and Ultrasound) segmentation tasks in terms of DSC and NSD. Scores are displayed as Median values (First quartile, Third quartile). Scores marked with * indicate statistically significant superiority over SAM, while those marked with † denote statistically significant superiority of MedSAM over the specialist U-Net model (p -value < 0.05).

Target	Modality	DSC (%)			NSD (%)		
		SAM	U-Net	MedSAM	SAM	U-Net	MedSAM
COVID-19	CXR	78.2 (70.6, 85.3)	91.1 (86.5, 94.7)*	92.9 (89.3, 95.6)*†	86.5 (80.8, 92.0)	97.2 (94.5, 98.9)*	98.3 (96.5, 99.3)*†
Heart	CXR	90.1 (80.5, 93.5)	95.6 (94.5, 96.2)*	96.8 (96.0, 97.3)*†	94.7 (85.9, 97.5)	98.7 (98.2, 99.2)*	99.4 (98.8, 99.7)*†
Lung	CXR	93.3 (90.1, 95.2)	98.2 (97.5, 98.5)*	99.1 (98.3, 99.4)*†	97.6 (94.6, 99.0)	100 (99.8, 100)*	100 (99.9, 100)*†
Pneumothorax	CXR	50.2 (38.1, 62.9)	76.3 (67.8, 83.2)*	81.5 (75.1, 86.8)*†	67.5 (54.3, 81.0)	92.6 (85.0, 96.7)*	96.4 (93.0, 98.2)*†
Tuberculosis	CXR	93.9 (89.7, 95.6)	96.8 (95.3, 97.4)*	96.9 (95.4, 97.6)*†	97.8 (93.7, 99.0)	99.7 (98.7, 99.8)*	99.6 (98.3, 99.9)*
Breast Cancer	Mammography	66.5 (46.5, 77.7)	78.4 (70.9, 85.4)*	83.3 (78.3, 87.5)*†	80.4 (60.6, 89.8)	91.8 (84.1, 96.0)*	94.8 (91.0, 97.3)*†
Diabetic Macular Edema	OCT	88.4 (83.1, 91.6)	91.5 (86.8, 95.1)*	95.0 (92.9, 96.3)*†	99.0 (97.1, 100)	99.9 (98.8, 100)*	100 (99.8, 100)*†
Benign Breast Cancer	Ultrasound	89.3 (81.9, 92.9)	93.5 (90.1, 94.8)*	94.0 (92.6, 95.2)*†	96.4 (88.6, 99.2)	98.8 (97.6, 99.5)*	99.1 (98.0, 99.7)*
Benign Thyroid Nodule	Ultrasound	88.1 (83.0, 91.9)	93.8 (91.1, 95.8)*	94.8 (92.8, 96.3)*†	93.7 (89.5, 96.6)	98.0 (96.3, 99.1)*	98.7 (97.8, 99.4)*†
Fetal Head	Ultrasound	84.9 (78.7, 88.9)	96.1 (95.0, 96.9)*	98.1 (97.3, 98.6)*†	91.2 (86.3, 94.7)	99.4 (98.6, 99.8)*	100 (99.8, 100)*†
Kidney	Ultrasound	88.0 (82.6, 92.5)	96.9 (95.4, 97.7)*	98.1 (97.3, 98.6)*†	92.5 (87.5, 96.5)	99.6 (98.8, 99.8)*	99.9 (99.7, 100)*†
Left Atrium	Ultrasound	44.7 (4.2, 81.6)	94.1 (92.3, 95.4)*	98.3 (97.7, 98.7)*†	68.3 (7.4, 92.5)	98.9 (98.1, 99.5)*	100 (100, 100)*†
Left Ventricle	Ultrasound	81.8 (73.0, 86.3)	95.2 (93.8, 96.2)*	98.4 (97.9, 98.6)*†	91.0 (85.8, 93.9)	99.2 (98.6, 99.6)*	100 (100, 100)*†
Malignant Breast Cancer	Ultrasound	82.9 (72.0, 87.6)	88.2 (85.6, 91.0)*	91.3 (88.2, 92.9)*†	91.4 (80.1, 92.4)	93.7 (92.0, 96.2)*	95.8 (93.9, 97.4)*†
Malignant Thyroid Nodule	Ultrasound	88.1 (82.1, 91.8)	91.3 (88.4, 93.7)*	93.1 (90.7, 94.7)*†	95.6 (91.4, 97.9)	98.1 (96.5, 99.2)*	98.9 (97.7, 99.5)*†
Nerve Cancer	Ultrasound	55.3 (40.1, 71.0)	89.5 (87.0, 91.8)*	90.6 (88.2, 92.3)*†	75.0 (58.8, 85.4)	97.2 (95.2, 98.6)*	97.7 (96.2, 98.7)*†
Pubic Symphysis	Ultrasound	75.7 (67.3, 80.2)	92.6 (90.4, 94.5)*	97.1 (95.6, 98.3)*†	87.4 (81.8, 90.9)	99.6 (98.7, 99.9)*	100 (100, 100)*†

TABLE 8

Internal validation results of SAM, specialist U-Net, and MedSAM on RGB images (Dermoscopy, Endoscopy, Fundus, and Pathology) segmentation tasks in terms of DSC and NSD. Scores are displayed as Median values (First quartile, Third quartile). Scores marked with * indicate statistically significant superiority over SAM, while those marked with † denote statistically significant superiority of MedSAM over the specialist U-Net model (p -value < 0.05).

Target	Modality	DSC (%)			NSD (%)		
		SAM	U-Net	MedSAM	SAM	U-Net	MedSAM
Skin Cancer	Dermoscopy	88.8 (79.9, 93.2)	95.1 (92.2, 97.0)*	95.2 (93.0, 96.7)*†	94.9 (85.9, 98.3)	99.4 (98.6, 99.8)*	99.1 (97.7, 99.7)*†
Cholecystectomy	Endoscopy	71.8 (62.7, 81.5)	93.9 (89.9, 96.0)*	95.5 (93.3, 96.9)*†	81.1 (72.1, 89.8)	98.8 (96.3, 99.5)*	99.5 (98.7, 99.8)*†
Gastrectomy	Endoscopy	81.5 (72.4, 89.0)	89.2 (82.6, 94.4)*	97.4 (95.5, 98.7)*†	90.6 (82.4, 96.7)	96.5 (91.1, 99.1)*	99.9 (99.5, 100)*†
Polyp	Endoscopy	94.1 (89.9, 96.3)	94.8 (91.4, 96.2)	98.4 (97.9, 98.9)*†	99.1 (96.8, 99.8)	99.6 (98.2, 99.9)	100 (100, 100)*†
Surgical Instrument	Endoscopy	87.0 (80.3, 90.1)	96.5 (94.9, 97.3)*	96.7 (95.6, 97.5)*	93.9 (87.7, 96.9)	99.8 (99.4, 99.9)*	99.8 (99.5, 99.9)*
Glaucoma Cup	Fundus	79.0 (73.2, 83.2)	80.2 (75.5, 86.5)	97.8 (96.5, 98.3)*†	90.2 (87.4, 94.6)	95.6 (90.1, 96.5)*	100 (100, 100)*†
Glaucoma Disc	Fundus	85.4 (82.3, 87.2)	95.7 (94.1, 96.4)*	98.3 (97.5, 98.6)*†	96.8 (93.6, 97.5)	100 (99.6, 100)*	100 (100, 100)*†
Non-Glaucoma Cup	Fundus	69.8 (63.4, 75.6)	81.3 (74.2, 88.0)*	96.2 (95.4, 97.3)*†	90.1 (85.1, 95.6)	97.0 (90.2, 99.8)*	100 (100, 100)*†
Non-Glaucoma Disc	Fundus	85.1 (76.0, 89.3)	94.4 (93.3, 95.4)*	98.3 (98.0, 98.8)*†	95.7 (90.7, 98.6)	100 (99.8, 100)*	100 (100, 100)*†
Glomeruli	Pathology	93.3 (90.8, 95.2)	97.1 (96.2, 97.8)*	98.5 (98.3, 98.7)*†	98.1 (96.5, 99.2)	99.8 (99.6, 100)*	100 (100, 100)*†
Intestine FTU	Pathology	89.0 (84.6, 92.7)	93.9 (92.5, 95.1)*	94.2 (92.5, 95.0)*	96.7 (93.2, 98.4)	99.2 (98.4, 99.5)*	99.1 (98.6, 99.5)*

TABLE 9

External validation results of SAM, specialist U-Net, and MedSAM on CT image segmentation tasks in terms of DSC and NSD. Scores are displayed as Median values (First quartile, Third quartile). Scores marked with * indicate statistically significant superiority over SAM, while those marked with † denote statistically significant superiority of MedSAM over the specialist U-Net model (p -value < 0.05).

Target	Modality	DSC (%)			NSD (%)		
		SAM	U-Net	MedSAM	SAM	U-Net	MedSAM
Aorta	CT	91.2 (89.7, 92.6)	85.3 (78.0, 89.4)*	93.1 (91.2, 94.7)*†	92.4 (89.9, 94.3)	81.6 (69.8, 90.7)*	98.7 (95.0, 99.6)*†
Esophagus	CT	68.0 (62.3, 73.9)	48.9 (41.5, 58.5)*	67.4 (59.6, 77.2)*†	81.8 (76.0, 87.1)	47.2 (37.3, 60.7)*	80.3 (71.9, 89.7)*†
Gallbladder	CT	78.9 (71.2, 86.0)	72.2 (62.1, 80.4)*	84.6 (76.7, 88.6)*†	88.6 (83.2, 93.3)	78.0 (70.2, 86.6)*	93.9 (87.5, 98.0)*†
Heart	CT	91.3 (90.2, 92.0)	90.7 (88.6, 91.9)	96.8 (96.7, 97.2)*†	88.8 (86.9, 90.9)	88.5 (84.0, 89.6)	99.5 (99.3, 99.8)*†
Inferior Vena Cava	CT	76.2 (71.0, 79.5)	63.0 (57.2, 70.6)*	79.5 (71.6, 84.0)*†	72.2 (67.9, 77.5)	52.6 (46.9, 61.2)*	78.8 (69.5, 86.6)*†
Left Kidney	CT	93.7 (92.6, 94.7)	93.0 (92.3, 93.8)*	93.7 (92.9, 94.4)*†	92.7 (90.6, 95.8)	94.1 (92.2, 96.2)*	94.9 (93.3, 96.8)*†
Left Lung	CT	91.5 (89.6, 94.3)	92.4 (90.8, 94.3)	96.6 (96.1, 97.2)*†	86.9 (83.4, 92.9)	88.9 (84.1, 92.7)	98.9 (98.1, 99.5)*†
Liver	CT	91.6 (88.4, 93.4)	94.3 (93.1, 95.0)*	94.8 (93.6, 95.6)*†	76.4 (70.4, 82.3)	90.2 (86.6, 92.7)*	93.4 (89.5, 95.5)*†
Liver Cancer	CT	76.6 (70.1, 85.1)	87.3 (82.4, 91.5)*	96.0 (94.7, 96.9)*†	67.7 (59.3, 77.1)	80.0 (75.6, 84.3)*	98.6 (97.1, 99.5)*†
Lung Cancer	CT	64.2 (53.5, 70.9)	55.8 (52.0, 70.3)*	77.8 (68.7, 90.2)*†	77.5 (74.4, 83.7)	68.1 (63.1, 79.4)*	89.5 (86.5, 98.0)*†
Nasopharynx Cancer	CT	37.0 (31.5, 44.4)	65.8 (61.5, 70.9)*	90.3 (87.8, 93.2)*†	71.0 (66.1, 75.5)	68.7 (65.5, 75.4)	97.9 (96.6, 99.7)*†
Pancreas	CT	72.8 (65.9, 77.0)	73.7 (69.5, 78.1)*	80.6 (77.4, 84.9)*†	77.0 (71.7, 81.3)	81.6 (75.4, 86.5)*	90.0 (85.0, 93.5)*†
Right Kidney	CT	93.5 (92.9, 94.4)	93.2 (92.4, 94.0)*	93.5 (93.0, 94.2)*†	92.2 (90.4, 94.3)	94.3 (92.1, 96.0)*	94.2 (92.6, 95.9)*
Right Lung	CT	92.6 (91.3, 94.5)	93.6 (92.1, 94.9)	97.4 (97.0, 97.7)*†	83.3 (80.6, 90.8)	89.9 (83.2, 93.0)*	99.2 (98.8, 99.6)*†
Spinal Cord	CT	47.5 (39.6, 56.2)	38.6 (35.5, 41.8)*	58.1 (55.8, 62.4)*†	58.5 (53.2, 63.9)	26.6 (21.6, 31.5)*	58.9 (50.3, 65.1)*
Spleen	CT	93.3 (89.6, 95.1)	94.0 (93.0, 94.8)*	94.5 (93.6, 95.2)*†	92.2 (84.5, 96.3)	96.8 (94.7, 98.2)*	98.7 (97.8, 99.2)*†
Stomach	CT	85.5 (80.5, 89.6)	90.1 (87.2, 92.3)*	93.1 (91.4, 94.5)*†	77.8 (71.6, 82.5)	85.8 (81.1, 91.2)*	94.8 (90.7, 97.5)*†

TABLE 10

External validation results of SAM, specialist U-Net, and MedSAM on MR image segmentation tasks in terms of DSC and NSD. Scores are displayed as Median values (First quartile, Third quartile). Scores marked with * indicate statistically significant superiority over SAM, while those marked with † denote statistically significant superiority of MedSAM over the specialist U-Net model (p -value < 0.05).

Target	Modality	DSC (%)			NSD (%)		
		SAM	U-Net	MedSAM	SAM	U-Net	MedSAM
A Carotid L	MR	57.8 (54.7, 62.3)	50.7 (47.9, 53.7)*	62.0 (59.5, 65.4)*†	70.0 (65.6, 73.1)	58.7 (53.3, 63.4)*	74.5 (68.8, 79.1)*†
A Carotid R	MR	61.0 (55.6, 64.2)	51.5 (48.6, 55.2)*	62.7 (59.3, 66.9)*†	72.0 (68.2, 75.1)	58.3 (56.4, 61.2)*	74.1 (69.3, 81.2)*†
Bone Mandible	MR	54.3 (50.2, 58.3)	22.6 (19.5, 25.9)*	69.7 (66.7, 73.7)*†	47.9 (45.7, 51.9)	35.1 (31.8, 38.6)*	60.0 (56.7, 63.9)*†
Brain Cancer	MR	85.4 (78.1, 90.4)	82.5 (76.7, 88.1)	94.5 (93.6, 96.1)*†	87.3 (75.7, 96.4)	91.6 (81.3, 94.2)	100 (99.8, 100)*†
Brainstem	MR	84.4 (83.3, 86.9)	88.6 (88.0, 89.3)*	96.0 (95.6, 96.2)*†	79.8 (77.6, 85.5)	88.1 (85.6, 89.7)*	99.7 (99.5, 99.9)*†
Cervical Cancer	MR	76.9 (70.1, 81.6)	81.6 (78.0, 86.0)*	89.7 (88.3, 90.9)*†	88.6 (84.9, 90.5)	92.8 (91.4, 94.2)	98.2 (97.2, 99.1)*†
Cricopharyngeus	MR	61.4 (55.5, 64.6)	61.1 (56.6, 64.0)	90.2 (83.1, 91.5)*†	68.5 (63.1, 73.3)	71.5 (68.8, 79.6)*	99.2 (97.0, 99.8)*†
Eye PL	MR	81.5 (78.5, 83.6)	82.7 (81.1, 83.8)*	94.1 (93.3, 94.7)*†	84.1 (78.4, 88.5)	86.6 (82.7, 89.9)*	99.6 (99.1, 99.8)*†
Eye PR	MR	81.9 (80.1, 83.8)	81.9 (79.4, 83.8)	94.0 (93.0, 94.4)*†	85.8 (80.5, 91.8)	85.7 (83.1, 88.2)	99.6 (99.3, 99.9)*†
Glnd Lacrimal L	MR	61.3 (58.6, 63.4)	51.5 (50.1, 56.2)*	64.0 (62.5, 66.7)*†	74.1 (70.7, 80.7)	59.3 (55.6, 66.9)*	79.7 (74.3, 86.1)*†
Glnd Lacrimal R	MR	59.9 (57.5, 66.9)	51.6 (48.7, 60.1)*	68.7 (60.5, 76.0)*†	79.8 (75.0, 84.3)	65.1 (58.3, 69.0)*	88.9 (79.7, 92.6)*†
Glnd Submand L	MR	77.9 (74.9, 81.1)	76.6 (74.3, 79.1)	91.3 (90.2, 91.9)*†	81.4 (76.9, 85.4)	80.2 (73.7, 83.3)*	98.7 (98.0, 99.3)*†
Glnd Submand R	MR	79.7 (74.3, 82.7)	77.2 (72.9, 79.6)*	90.9 (89.2, 92.5)*†	85.8 (78.7, 88.6)	81.0 (74.1, 83.8)*	98.9 (98.4, 99.4)*†
Glottis	MR	30.1 (14.4, 42.7)	44.2 (35.6, 49.6)*	85.0 (79.1, 88.6)*†	61.6 (53.7, 67.7)	66.6 (62.0, 71.1)*	97.9 (95.6, 99.5)*†
Heart	MR	52.1 (36.3, 66.8)	91.3 (88.5, 93.1)*	96.3 (95.2, 96.9)*†	75.0 (62.8, 84.5)	99.2 (98.2, 99.7)*	100 (99.9, 100)*†
Larynx SG	MR	54.0 (46.9, 60.6)	60.4 (56.3, 64.5)*	88.2 (85.2, 90.4)*†	71.0 (65.1, 75.3)	72.4 (65.6, 78.2)	97.4 (95.9, 98.8)*†
Left Kidney	T1-Inphase	89.2 (86.0, 92.1)	91.8 (88.5, 92.3)	95.0 (93.4, 96.5)*†	94.8 (92.0, 97.8)	96.6 (94.3, 97.7)	99.9 (99.6, 100)*†
Left Kidney	T1-Outphase	89.7 (86.6, 91.9)	89.6 (88.0, 91.0)	93.3 (92.9, 94.3)*†	96.0 (92.4, 98.3)	96.0 (94.2, 97.3)	99.6 (99.3, 99.9)*†
Left Kidney	T2	92.8 (91.2, 94.3)	94.1 (93.2, 94.8)*	96.1 (94.7, 96.7)*†	97.6 (94.3, 98.3)	98.4 (97.3, 99.3)*	99.6 (99.4, 99.9)*†
Lips	MR	58.4 (50.9, 66.5)	60.3 (50.6, 67.1)	86.9 (83.9, 89.3)*†	69.2 (63.6, 72.5)	62.6 (59.3, 68.2)*	96.3 (92.8, 96.8)*†
Liver	T1-Inphase	89.0 (88.0, 90.0)	92.1 (90.0, 92.8)*	95.6 (95.3, 96.0)*†	89.5 (82.9, 91.4)	92.2 (90.9, 95.4)*	98.7 (98.2, 99.2)*†
Liver	T1-Outphase	90.9 (89.7, 92.6)	93.6 (92.3, 94.0)*	95.0 (94.5, 96.1)*†	92.9 (88.4, 95.4)	96.2 (94.7, 97.4)*	98.7 (97.5, 99.1)*†
Liver	T2	90.4 (89.0, 91.7)	92.4 (91.3, 93.2)*	96.4 (95.2, 96.7)*†	91.1 (88.6, 93.2)	94.6 (91.7, 96.1)*	98.9 (98.4, 99.7)*†
OpticNrv L	MR	39.5 (34.2, 50.0)	41.9 (38.3, 54.3)*	61.3 (59.6, 66.3)*†	56.3 (52.3, 72.4)	55.5 (51.7, 70.6)	83.9 (79.2, 87.1)*†
OpticNrv R	MR	43.3 (29.7, 49.6)	41.3 (35.0, 47.1)	70.3 (64.6, 77.5)*†	59.4 (47.2, 66.3)	53.2 (47.4, 59.0)	92.4 (85.3, 94.0)*†
Parotid L	MR	72.7 (63.5, 77.5)	79.1 (77.5, 81.6)*	91.7 (90.6, 92.4)*†	67.5 (59.8, 75.0)	74.7 (68.6, 79.8)*	96.8 (95.4, 97.6)*†
Parotid R	MR	71.4 (67.5, 75.3)	77.3 (75.0, 79.3)*	91.6 (90.5, 92.3)*†	69.4 (63.3, 74.6)	69.4 (65.5, 75.0)	96.8 (95.8, 97.7)*†
Right Kidney	T1-Inphase	91.0 (87.6, 93.0)	91.0 (88.4, 92.6)	94.8 (94.1, 96.5)*†	97.2 (94.8, 98.9)	96.3 (94.1, 98.2)	99.9 (99.8, 100)*†
Right Kidney	T1-Outphase	89.4 (85.7, 91.4)	89.0 (87.5, 90.6)	93.6 (92.5, 94.9)*†	94.3 (92.6, 97.9)	95.2 (92.4, 96.9)	99.8 (99.5, 99.9)*†
Right Kidney	T2	94.0 (93.2, 94.4)	93.9 (92.2, 95.5)	95.9 (94.6, 97.1)*†	98.6 (97.7, 99.0)	97.1 (95.4, 98.0)*	99.9 (99.0, 100)*†
SpinalCord	MR	55.9 (54.3, 58.7)	64.3 (60.7, 67.7)*	77.4 (75.3, 80.6)*†	40.8 (35.9, 45.0)	57.9 (52.6, 63.0)*	85.5 (79.4, 91.1)*†
Spleen	T1-Inphase	86.4 (83.7, 88.5)	87.8 (86.1, 89.7)	94.3 (92.9, 95.0)*†	91.5 (85.3, 95.7)	93.6 (89.8, 94.7)	99.9 (99.7, 99.9)*†
Spleen	T1-Outphase	88.8 (86.3, 90.4)	88.6 (86.8, 91.4)	93.0 (91.6, 93.9)*†	96.5 (93.4, 98.8)	96.0 (88.8, 98.6)	99.7 (99.4, 99.9)*†
Spleen	T2	94.4 (92.9, 95.1)	91.9 (90.1, 94.7)	97.0 (95.5, 97.2)*†	99.7 (97.8, 99.9)	94.2 (89.7, 99.2)*	100 (99.9, 100)*†

TABLE 11

External validation results of SAM, specialist U-Net, and MedSAM on Chest X-Ray (CXR), Dermoscopy, Endoscopy, Fundus, and Pathology image segmentation tasks in terms of DSC and NSD. Scores are displayed as Median values (First quartile, Third quartile). Scores marked with * indicate statistically significant superiority over SAM, while those marked with † denote statistically significant superiority of MedSAM over the specialist U-Net model (p -value < 0.05).

Target	Modality	DSC (%)			NSD (%)		
		SAM	U-Net	MedSAM	SAM	U-Net	MedSAM
COVID-19	CXR	91.4 (88.2, 94.1)	97.7 (97.1, 98.2)*	98.9 (98.3, 99.2)*†	96.1 (93.3, 98.3)	99.9 (99.8, 100)*	100 (99.9, 100)*†
Lung Opacity	CXR	89.0 (84.7, 92.0)	97.2 (96.1, 97.8)*	98.3 (97.6, 98.8)*†	94.6 (90.9, 97.0)	99.8 (99.4, 99.9)*	99.9 (99.7, 100)*†
Viral Pneumonia	CXR	89.2 (82.7, 92.3)	96.8 (95.6, 97.3)*	98.4 (97.5, 98.8)*†	94.9 (88.9, 97.2)	99.8 (99.3, 99.9)*	99.9 (99.7, 100)*†
Skin Cancer	Dermoscopy	94.1 (91.6, 96.0)	90.5 (86.7, 95.5)*	96.9 (96.3, 97.4)*†	99.4 (96.6, 99.8)	97.7 (95.9, 99.4)	99.9 (99.6, 100)*†
Polyp	Endoscopy	93.8 (89.8, 95.8)	95.9 (93.9, 97.0)*	98.0 (97.6, 98.4)*†	98.2 (95.8, 99.4)	99.3 (98.1, 99.7)*	99.9 (99.8, 100)*†
Optic Disc	Fundus	82.8 (73.4, 89.3)	93.0 (85.1, 95.5)*	97.4 (96.0, 98.1)*†	95.1 (89.1, 98.2)	99.2 (97.0, 99.8)*	100 (100, 100)*†
Colon Gland	Pathology	81.3 (75.4, 84.8)	96.0 (95.2, 96.7)*	95.9 (95.4, 96.7)*	89.1 (85.4, 92.1)	99.6 (99.2, 99.8)*	99.6 (99.3, 99.8)*

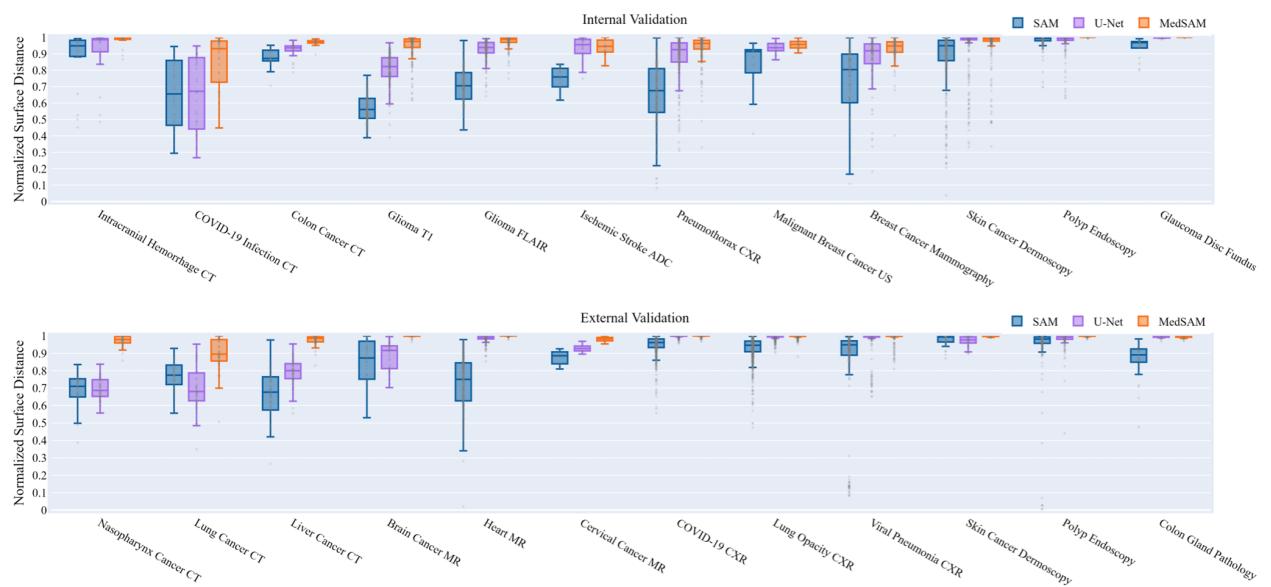


Fig. 2. Normalized Surface Distance for 12 representative segmentation tasks as part of the internal validation, and a further 12 tasks as part of the external validation (corresponding to tasks presented in Fig. 2 of the main text).



Fig. 3. Dice Similarity Coefficient of segmentation tasks in internal validation.

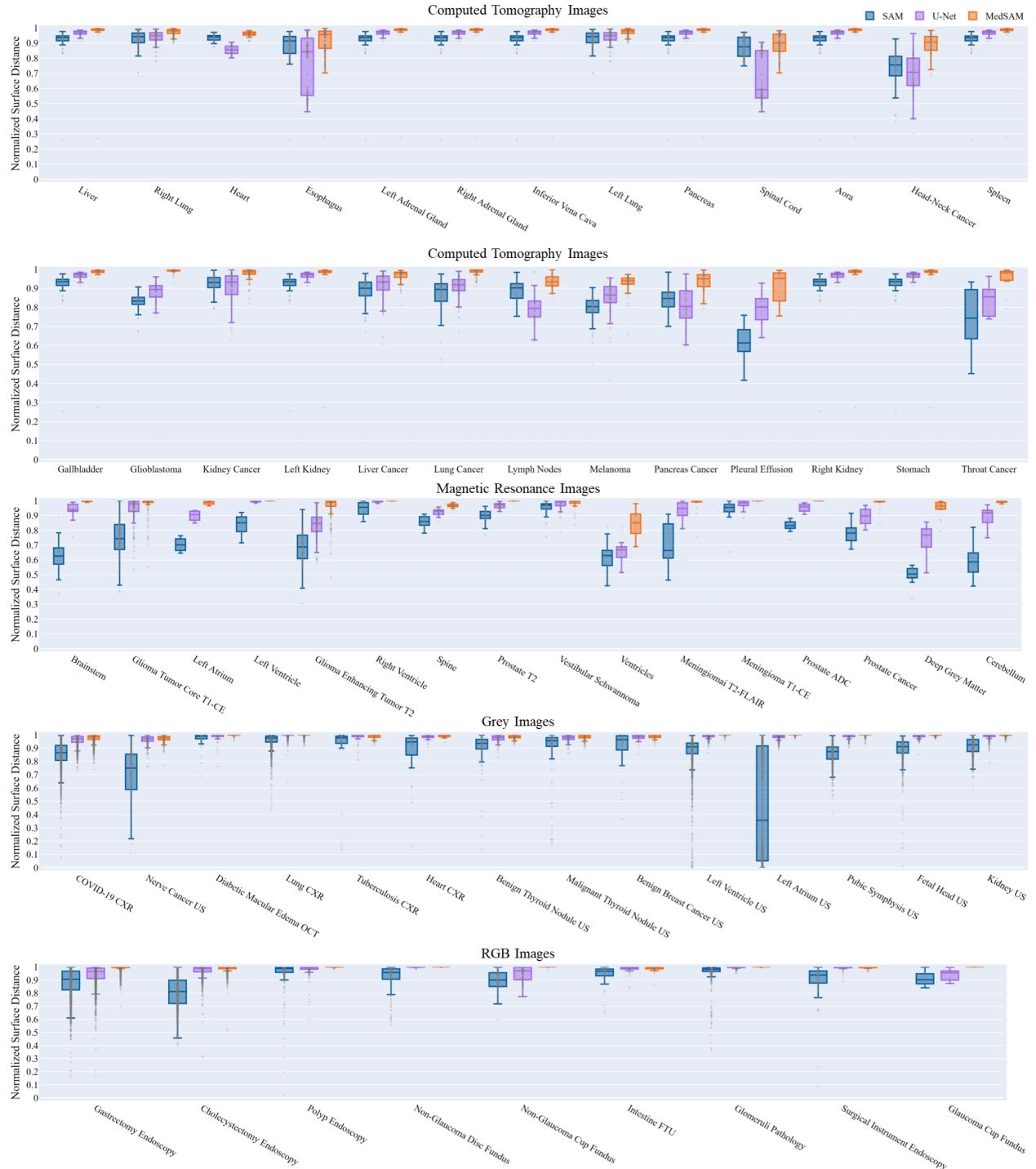


Fig. 4. Normalized surface distance of segmentation tasks in internal validation.

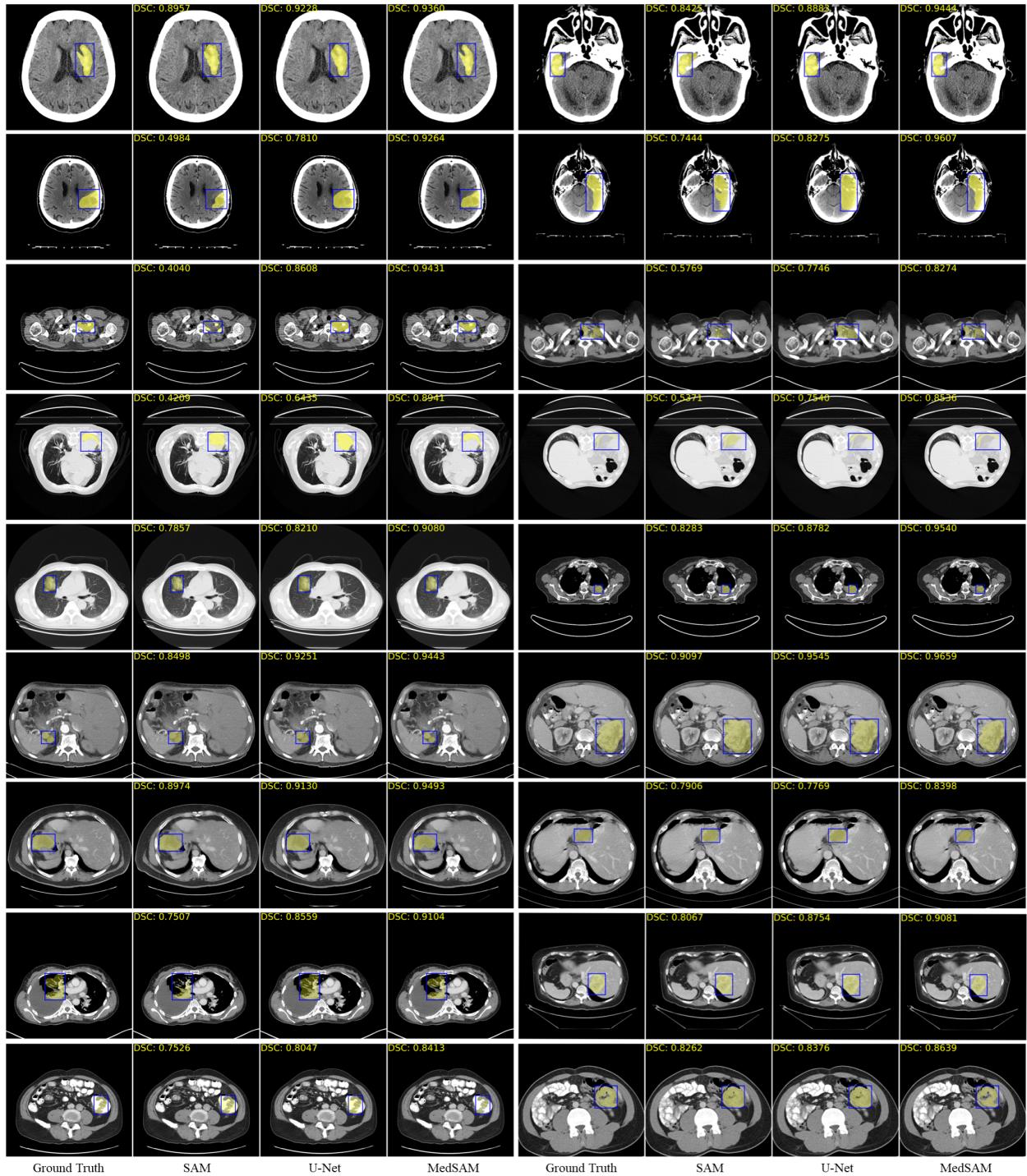


Fig. 5. Segmentation examples of CT images in internal validation.

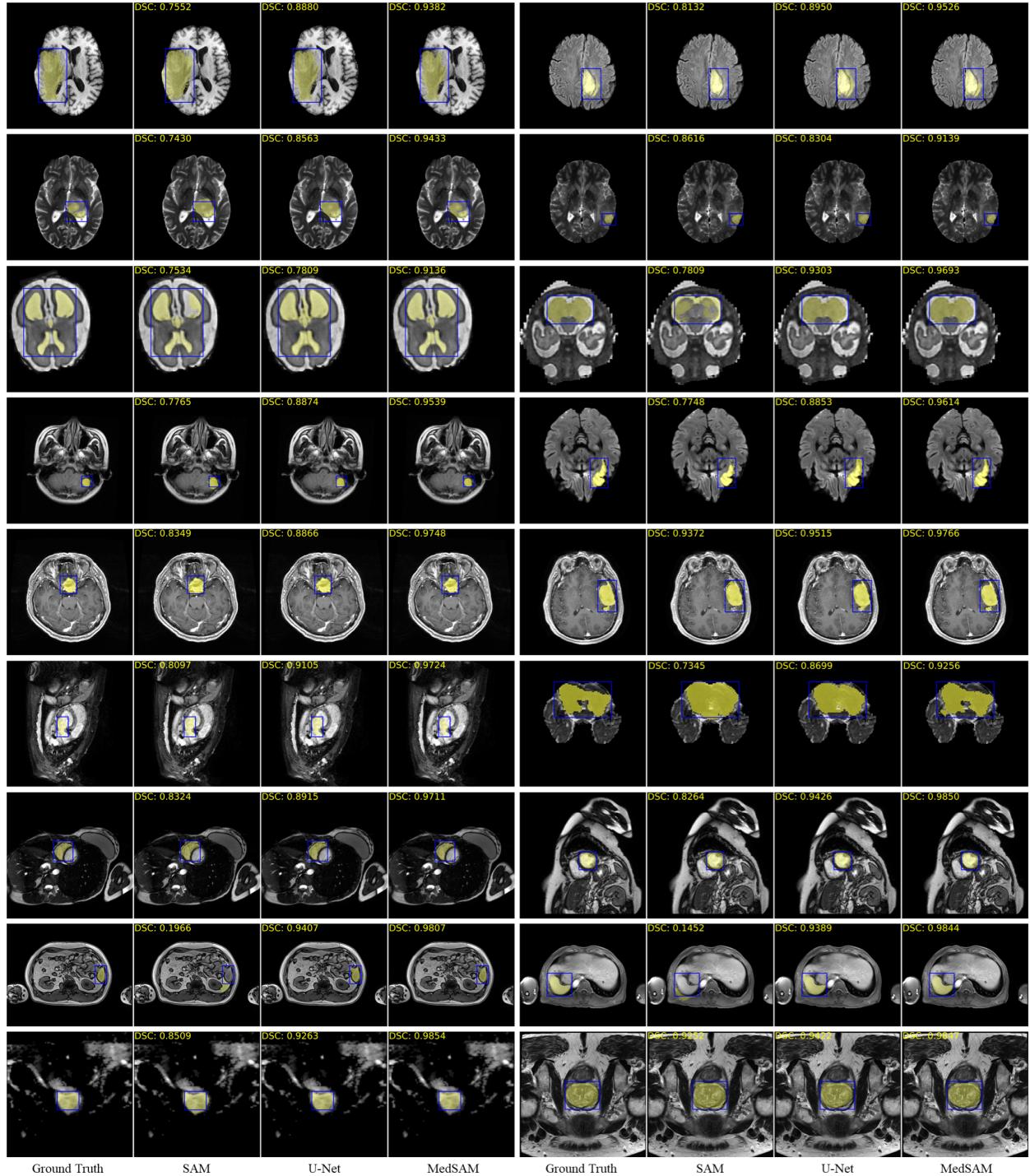


Fig. 6. Segmentation examples of MR images in internal validation.

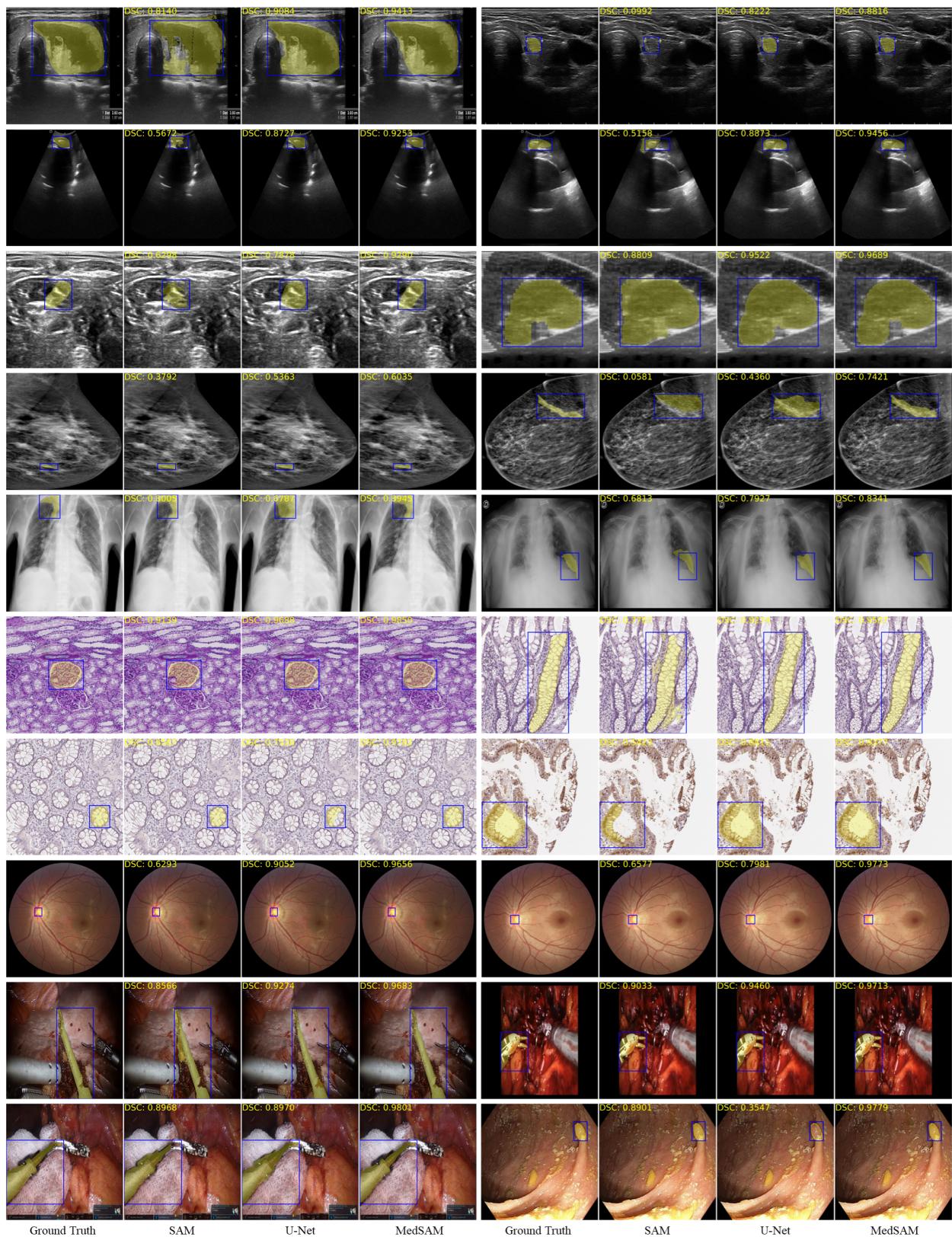


Fig. 7. Segmentation example of ultrasound, chest X-ray, pathology, fundus, and endoscopy images in internal validation.



Fig. 8. Dice similarity coefficient of segmentation tasks in external validation.

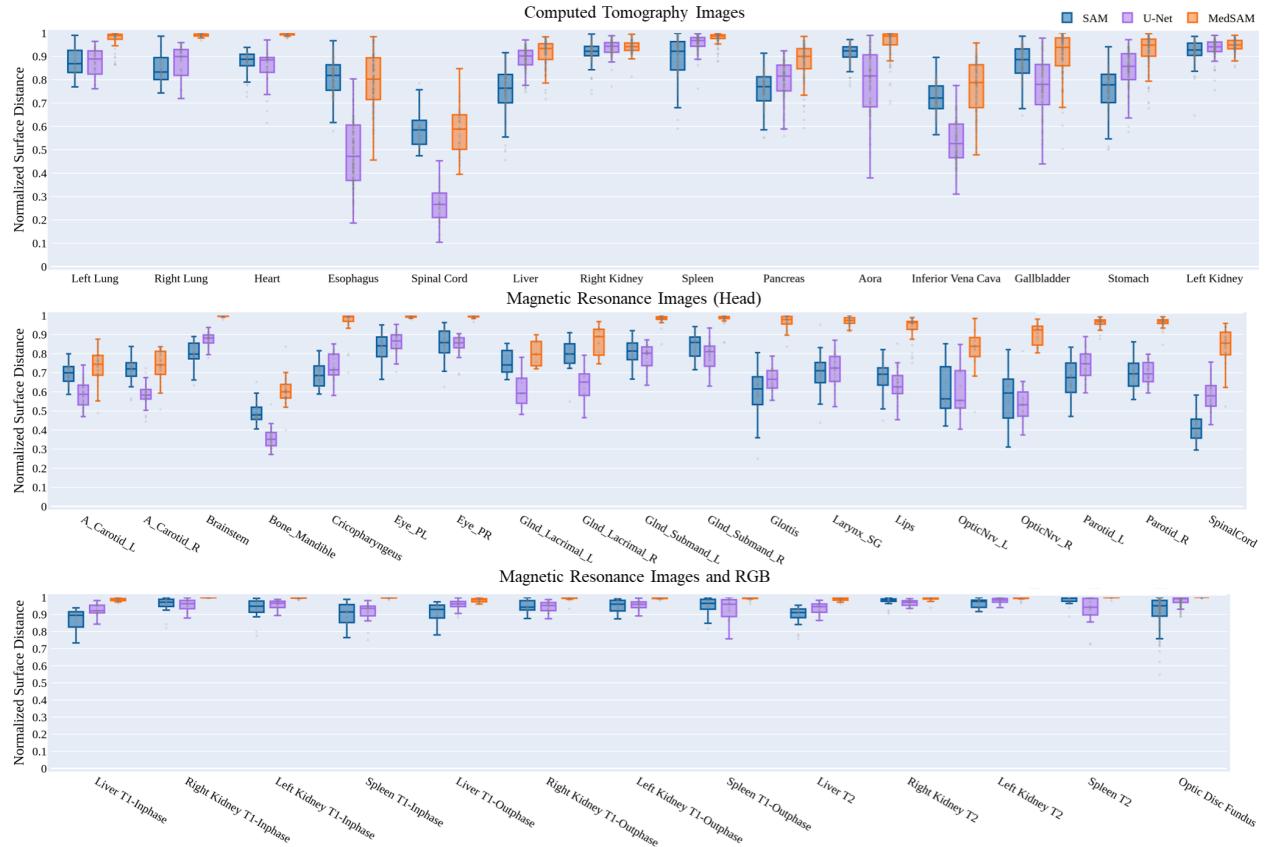


Fig. 9. Normalized surface distance of segmentation tasks in external validation.

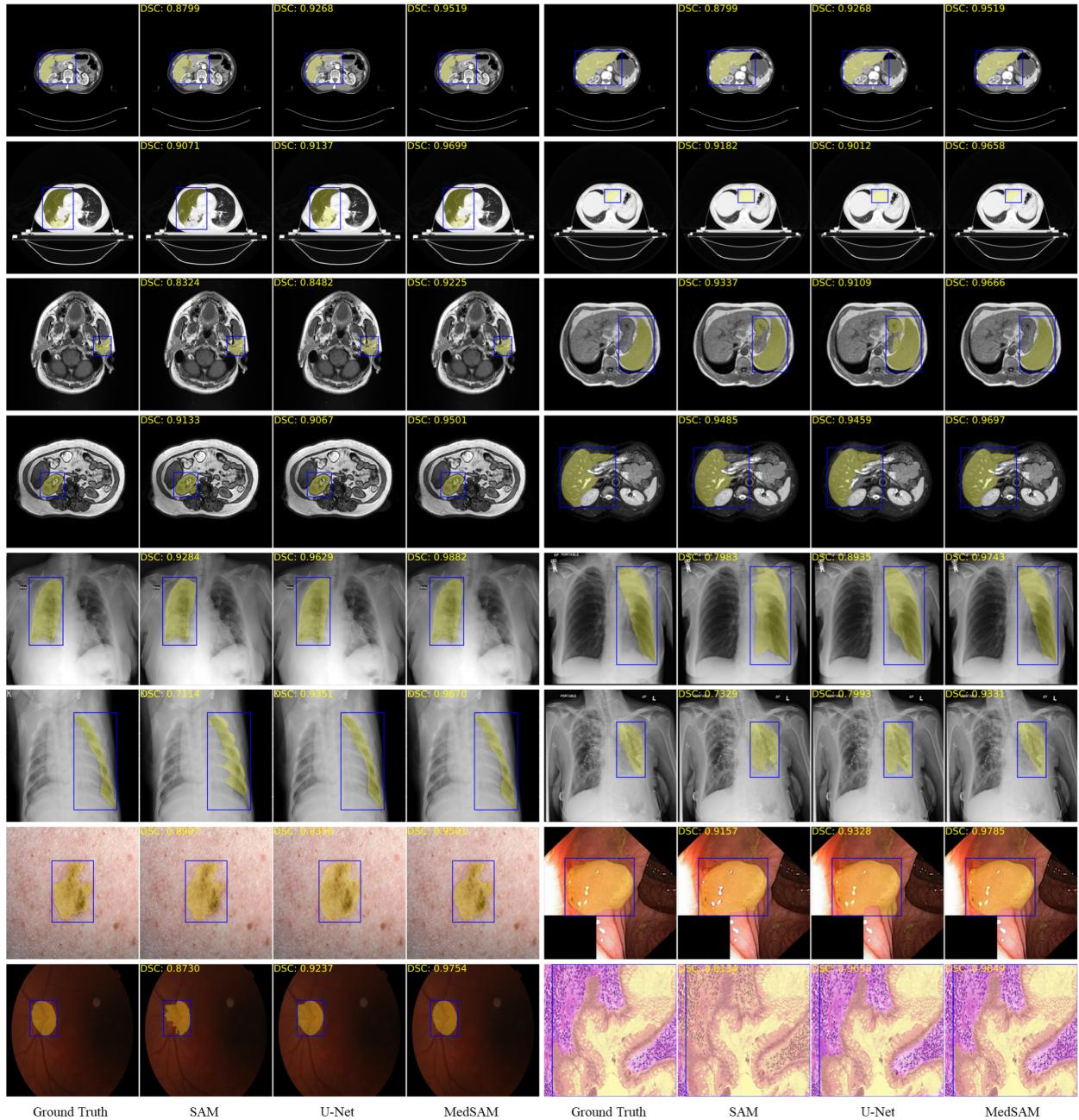


Fig. 10. Segmentation examples in external validation.

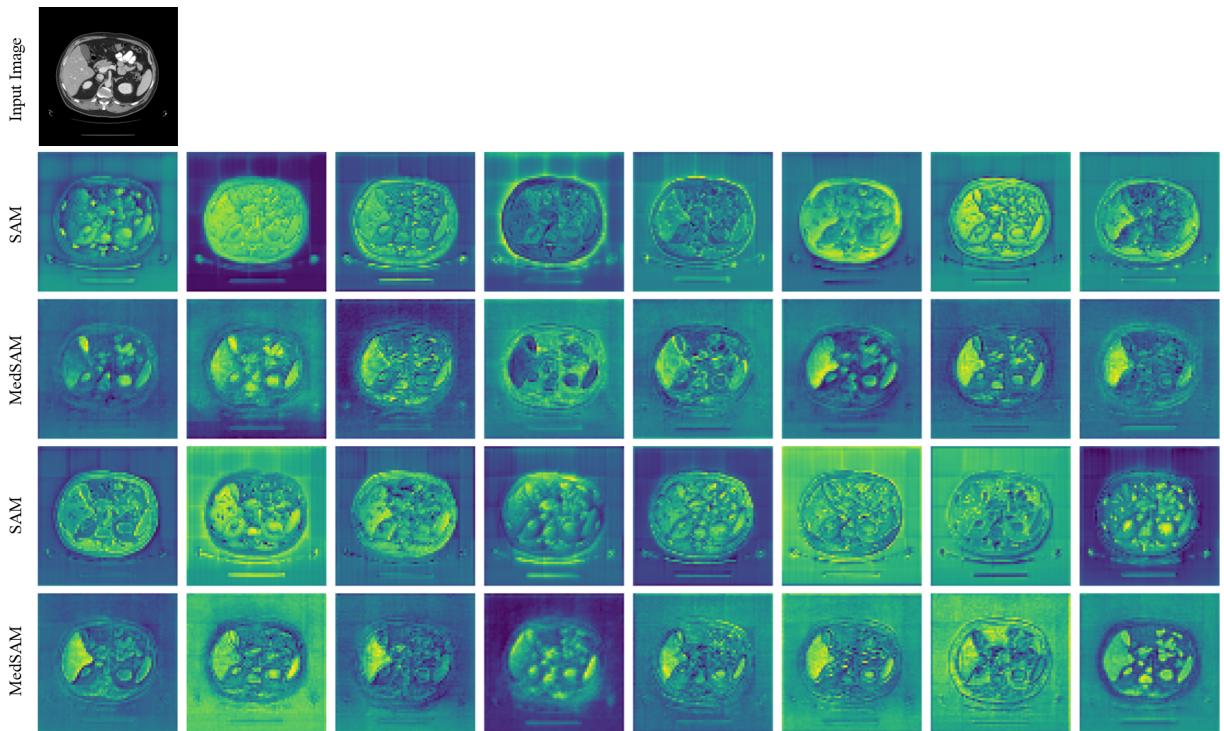


Fig. 11. Visualization and comparison of the saliency maps for image embeddings between SAM and MedSAM. MedSAM's features exhibited a greater abundance of semantic information, specifically pertaining to highly relevant anatomical structures.

TABLE 12
Data availability (CT and MR images).

Dataset	Modality	Download Link
AbdomenCT-1K	CT	https://github.com/JunMa11/AbdomenCT-1K
AMOS-CT	PET-CT	https://amos22.grand-challenge.org/
AutoPET	CT	https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=93258287
COVID-19 Seg. Challenge	CT	https://covid-segmentation.grand-challenge.org/Data/
COVID-19-CT-Seg	CT	https://github.com/JunMa11/COVID-19-CT-Seg-Benchmark
GLIS-RT	CT	https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=95224486
HCC-TACE-Seg	CT	https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70230229
HECKTOR	PET-CT	https://hecktor.grand-challenge.org/Overview/
INSTANCE	CT	https://instance.grand-challenge.org/
KiPA	CT	https://kipa22.grand-challenge.org/
KiTS	CT	https://kits-challenge.org/kits23/
Lymph Nodes	CT	https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=19726546
MMWHS-CT	CT	https://zmiclab.github.io/zxh/0/mmwhs/
MSD-Colon Tumor	CT	http://medicaldecathlon.com/
MSD-Hepatic Tumor	CT	http://medicaldecathlon.com/
MSD-Lung Tumor	CT	http://medicaldecathlon.com/
MSD-Pancreas	CT	http://medicaldecathlon.com/
MSD-Spleen	CT	http://medicaldecathlon.com/
NSCLC Pleural Effusion	CT	https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=68551327
NSCLC Radiogenomics	CT	https://wiki.cancerimagingarchive.net/display/Public/NSCLC+Radiogenomics
ORG	CT	https://www.nature.com/articles/s41597-020-00715-8
SegTHOR	CT	https://competitions.codalab.org/competitions/21145
StructSeg	CT	https://structseg2019.grand-challenge.org/Dataset/
TotalSegmentator	CT	https://zenodo.org/record/6802614
WORD	CT	https://github.com/HilLab-git/WORD
ACDC	MR	https://humanheart-project.creatis.insa-lyon.fr/database/
AMOS-MR	MR	https://amos22.grand-challenge.org/Dataset/
ATLAS R2.0	MR	https://atlas.grand-challenge.org/
Brain Tumor Dataset Figshare	MR	https://www.kaggle.com/datasets/ashkhagan/figshare-brain-tumor-dataset
Brain TR-GammakKnife	MR	https://doi.org/10.7937/xb6d-py67
BraTS	MR	http://braintumorsegmentation.org/
CC-Tumor Heterogeneity	MR	https://doi.org/10.7937/ERZ5-QZ59
CHAOS	MR	https://chaos.grand-challenge.org/
crossMoDA	MR	https://crossmoda-challenge.ml/
FeTA	MR	https://feta.grand-challenge.org/
HaN-Seg	MR	https://zenodo.org/record/
ISLES	MR	http://www.isles-challenge.org/
I2CVB	MR	https://i2cvb.github.io/
Meningioma-SEG-CLASS	MR	https://doi.org/10.7937/0TKV-1A36
MMs	MR	https://www.ub.edu/mnms-2/
MMWHS-MR	MR	https://zmiclab.github.io/zxh/0/mmwhs/
MSD-Heart	MR	http://medicaldecathlon.com/
MSD-Prostate	MR	http://medicaldecathlon.com/#tasks
NCI-ISBI	MR	http://dx.doi.org/10.7937/K9/TCIA.2015.zF0vlOPv
PI-CAI	MR	http://github.com/DIAGNijmegen/picai_labels
PPMI	MR	https://www.ppmi-info.org/access-data-specimens/download-data
PROMISE	MR	https://promise12.grand-challenge.org/Details/
Qin-Prostate-Repeatability	MR	http://doi.org/10.7937/K9/TCIA.2018.MR1CKGND
QUBIQ	MR	https://qubiq21.grand-challenge.org/
Spine	MR	https://www.cg.informatik.uni-siegen.de/en/spine-segmentation-and-analysis
WMH	MR	https://wmh.isi.uu.nl/

TABLE 13

Data availability (Chest X-ray (CXR), Dermoscopy, Endoscopy, Fundus, Mammography, Optical Coherence Tomography (OCT), Pathology, Ultrasound images).

Dataset	Modality	Download Link
Chest Xray Masks and Labels	CXR	https://www.kaggle.com/datasets/nikhilpandey360/chest-xray-masks-and-labels
Chest X-Ray (Pneumothorax)	CXR	https://www.kaggle.com/datasets/vbookshelf/pneumothorax-chest-xray-images-and-masks
COVID-19 Radiography	CXR	https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database
COVID-QU-Ex	CXR	https://www.kaggle.com/datasets/anasmohammedtahir/covidqu
JSRT	CXR	http://imgcom.jsrt.or.jp/minisrtdb/
Lung	CXR	https://doi.org/10.2214/ajr.174.1.1740071
QaTa-COV19	CXR	https://www.kaggle.com/datasets/aysendegerli/qatacov19-dataset
ISIC	Dermoscopy	https://challenge.isic-archive.com/data/
UWaterloo Skin Cancer	Dermoscopy	https://uwaterloo.ca/vision-image-processing-lab/research-demos/skin-cancer-detection
BKAI-IGH NeoPolyp	Endoscopy	https://www.kaggle.com/competitions/bkai-igh-neopolyp/data
CholecSeg8k	Endoscopy	https://www.kaggle.com/datasets/newslab/cholecseg8k
Kvasir	Endoscopy	https://datasets.simula.no/kvasir/
m2caiSeg	Endoscopy	https://www.kaggle.com/datasets/salmanmaq/m2caiseg
PolypGen	Endoscopy	https://www.synapse.org/#!Synapse:syn45200214
RobTool	Endoscopy	https://www.synapse.org/#!Synapse:syn22427422
sisvse	Endoscopy	https://www.kaggle.com/datasets/yjh4374/sisvse-dataset
IDRiD	Fundus	https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid
PAPILA	Fundus	https://figshare.com/articles/dataset/PAPILA/14798004/1
REFUGE	Fundus	https://refuge.grand-challenge.org/
CDD-CESM	Mammography	https://doi.org/10.7937/29kw-ae92
Intraretinal Cystoid Fluid	OCT	https://www.kaggle.com/datasets/zeeshanahmed13/intraretinal-cystoid-fluid
OCT Images (DME)	OCT	https://www.kaggle.com/datasets/paultimothymooney/chiu-2015
GlaS@MIČCAI2015	Pathology	https://warwick.ac.uk/fac/cross_fac/tia/data/glascontest/download/
HuBMAP HPA	Pathology	https://www.kaggle.com/competitions/hubmap-organ-segmentation/
HuBMAP Hacking the Kidney	Pathology	https://www.kaggle.com/competitions/hubmap-kidney-segmentation/data?select=test
AbdomenUS	Ultrasound	https://www.kaggle.com/datasets/ignaciiorlando/ussimandsegm
Breast Cancer	Ultrasound	https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset
CAMUS	Ultrasound	https://www.creatis.insa-lyon.fr/Challenge/camus/
CT2USforKidneySeg	Ultrasound	https://www.kaggle.com/datasets/siatsyx/ct2usforkidneyseg
FH-PS-AOP	Ultrasound	https://ps-fh-aop-2023.grand-challenge.org/
HC	Ultrasound	https://hc18.grand-challenge.org/
TN-SCUI	Ultrasound	https://tn-scui2020.grand-challenge.org/
Nerve	Ultrasound	https://www.kaggle.com/competitions/ultrasound-nerve-segmentation/data

REFERENCES

- [1] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.
- [2] G. Wang, M. A. Zuluaga, W. Li, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin *et al.*, "Deepigeos: a deep interactive geodesic framework for medical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1559–1572, 2018.
- [3] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin *et al.*, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1562–1573, 2018.
- [4] T. Zhou, L. Li, G. Bredell, J. Li, J. Ünkelbach, and E. Konukoglu, "Volumetric memory network for interactive medical image segmentation," *Medical Image Analysis*, vol. 83, p. 102599, 2023.
- [5] X. Luo, G. Wang, T. Song, J. Zhang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, and S. Zhang, "Mideepseg: Minimally interactive segmentation of unseen objects from medical images using deep learning," *Medical image analysis*, vol. 72, p. 102102, 2021.
- [6] R. Deng, C. Cui, Q. Liu, T. Yao, L. W. Remedios, S. Bao, B. A. Landman, L. E. Wheless, L. A. Coburn, K. T. Wilson *et al.*, "Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging," *arXiv preprint arXiv:2304.04155*, 2023.
- [7] C. Hu and X. Li, "When sam meets medical images: An investigation of segment anything model (sam) on multi-phase liver tumor segmentation," *arXiv preprint arXiv:2304.08506*, 2023.
- [8] S. He, R. Bao, J. Li, P. E. Grant, and Y. Ou, "Accuracy of segment-anything model (sam) in medical image segmentation tasks," *arXiv preprint arXiv:2304.09324*, 2023.
- [9] S. Roy, T. Wald, G. Koehler, M. R. Rokuss, N. Disch, J. Holzschuh, D. Zimmerer, and K. H. Maier-Hein, "Sam. md: Zero-shot medical image segmentation capabilities of the segment anything model," *arXiv preprint arXiv:2304.05396*, 2023.
- [10] T. Zhou, Y. Zhang, Y. Zhou, Y. Wu, and C. Gong, "Can sam segment polyps?" *arXiv preprint arXiv:2304.07583*, 2023.
- [11] S. Mohapatra, A. Gosai, and G. Schlaug, "Sam vs bet: A comparative study for brain extraction and segmentation of magnetic resonance images using deep learning," *arXiv preprint arXiv:2304.04738*, 2023.
- [12] A. M. Maciej, D. Haoyu, G. Hanxue, Y. Jichen, K. Nicholas, and Z. Yixin, "Segment anything model for medical image analysis: an experimental study," *arXiv preprint arXiv:2304.10517*, 2023.
- [13] Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, R. Chen, J. Yu, J. Chen, C. Chen *et al.*, "Segment anything model for medical images?" *arXiv preprint arXiv:2304.14660*, 2023.
- [14] J. Chen and X. Bai, "Learning to" segment anything" in thermal infrared images through knowledge distillation with a large scale dataset satir," *arXiv preprint arXiv:2304.07969*, 2023.
- [15] L. Tang, H. Xiao, and B. Li, "Can sam segment anything? when sam meets camouflaged object detection," *arXiv preprint arXiv:2304.04709*, 2023.
- [16] G.-P. Ji, D.-P. Fan, P. Xu, M.-M. Cheng, B. Zhou, and L. Van Gool, "Sam struggles in concealed scenes—empirical study on" segment anything"," *arXiv preprint arXiv:2304.06022*, 2023.
- [17] W. Ji, J. Li, Q. Bi, W. Li, and L. Cheng, "Segment anything is not always perfect: An investigation of sam on different real-world applications," *arXiv preprint arXiv:2304.05750*, 2023.
- [18] T. Chen, L. Zhu, C. Ding, R. Cao, S. Zhang, Y. Wang, Z. Li, L. Sun, P. Mao, and Y. Zang, "Sam fails to segment anything?—sam-adapter: Adapting sam in underperformed scenes: Camouflage, shadow, and more," *arXiv preprint arXiv:2304.09148*, 2023.
- [19] J. Wu, R. Fu, H. Fang, Y. Liu, Z. Wang, Y. Xu, Y. Jin, and T. Arbel, "Medical sam adapter: Adapting segment anything model for medical image segmentation," *arXiv preprint arXiv:2304.12620*, 2023.
- [20] K. Zhang and D. Liu, "Customized segment anything model for medical image segmentation," *arXiv preprint arXiv:2304.13785*, 2023.
- [21] J. Ma, Y. Zhang, S. Gu, C. Zhu, C. Ge, Y. Zhang, X. An, C. Wang, Q. Wang, X. Liu, S. Cao, Q. Zhang, S. Liu, Y. Wang, Y. Li, J. He, and X. Yang, "Abdomenct-1k: Is abdominal organ segmentation a solved problem?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [22] J. Ma, Y. Zhang, S. Gu, X. An, Z. Wang, C. Ge, C. Wang, F. Zhang, Y. Wang, Y. Xu, S. Gou, F. Thaler, C. Payer, D. Štern, E. G. Henderson, D. M. McSweeney, A. Green, P. Jackson, L. McIntosh, Q.-C. Nguyen, A. Qayyum, P.-H. Conze, Z. Huang, Z. Zhou, D.-P. Fan, H. Xiong, G. Dong, Q. Zhu, J. He, and X. Yang, "Fast and low-gpu-memory abdomen ct organ segmentation: The flare challenge," *Medical Image Analysis*, vol. 82, p. 102616, 2022.
- [23] Y. Ji, H. Bai, J. Yang, C. Ge, Y. Zhu, R. Zhang, Z. Li, L. Zhang, W. Ma, X. Wan *et al.*, "Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation," *arXiv preprint arXiv:2206.08023*, 2022.
- [24] S. Gatidis, T. Hepp, M. Früh, C. La Fougère, K. Nikolaou, C. Pfannenberg, B. Schölkopf, T. Küstner, C. Cyran, and D. Rubin, "A whole-body fdg-pet/ct dataset with manually annotated tumor lesions," *Scientific Data*, vol. 9, no. 1, p. 601, 2022.
- [25] H. R. Roth, Z. Xu, C. Tor-Díez, R. S. Jacob, J. Zember, J. Molto, W. Li, S. Xu, B. Turkbey, E. Turkbey *et al.*, "Rapid artificial intelligence solutions in a pandemic—the covid-19-20 lung ct lesion segmentation challenge," *Medical Image Analysis*, vol. 82, p. 102605, 2022.
- [26] P. An, S. Xu, S. Harmon, E. B. Turkbey, T. H. Sanford, A. Amalou, M. Kassim, N. Varble, M. Blain, V. Anderson *et al.*, "Ct images in covid-19 [data set]," *The Cancer Imaging Archive*, vol. 10, 2020.
- [27] J. Ma, Y. Wang, X. An, C. Ge, Z. Yu, J. Chen, Q. Zhu, G. Dong, J. He, Z. He, T. Cao, Y. Zhu, Z. Nie, and X. Yang, "Towards data-efficient learning: A benchmark for covid-19 ct lung and infection segmentation," *Medical Physics*, vol. 48, no. 3, pp. 1197–1210, 2021.
- [28] N. Shusharina, T. Bortfeld, C. Cardenas, B. De, K. Diao, S. Hernandez, Y. Liu, S. Maroongroge, J. Söderberg, and M. Soliman, "Cross-modality brain structures image segmentation for the radiotherapy target definition and plan optimization," in *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data: MICCAI 2020 Challenges, ABCs 2020, L2R 2020, TN-SCUI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings* 23, 2021, pp. 3–15.
- [29] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.
- [30] A. Morshid, K. M. Elsayes, A. M. Khalaf, M. M. Elmohr, J. Yu, A. O. Kaseb, M. Hassan, A. Mahvash, Z. Wang, J. D. Hazle, and D. Fuentes, "A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization," *Radiology: Artificial Intelligence*, vol. 1, no. 5, p. e180021, 2019.
- [31] V. Oreiller, V. Andrearczyk, M. Jreige, S. Bougħdad, H. Elhalawani, J. Castelli, M. Vallières, S. Zhu, J. Xie, Y. Peng *et al.*, "Head and neck tumor segmentation in pet/ct: the hecktor challenge," *Medical Image Analysis*, vol. 77, p. 102336, 2022.
- [32] X. Li, G. Luo, W. Wang, K. Wang, Y. Gao, and S. Li, "Hematoma expansion context guided intracranial hemorrhage segmentation and uncertainty estimation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 1140–1151, 2021.
- [33] Y. He, G. Yang, J. Yang, R. Ge, Y. Kong, X. Zhu, S. Zhang, P. Shao, H. Shu, J.-L. Dillenseger *et al.*, "Meta grayscale adaptive network for 3d integrated renal structures segmentation," *Medical Image Analysis*, vol. 71, p. 102055, 2021.
- [34] Y. He, G. Yang, J. Yang, Y. Chen, Y. Kong, J. Wu, L. Tang, X. Zhu, J.-L. Dillenseger, P. Shao *et al.*, "Dense biased networks with deep prior anatomy and hard region adaptation: Semi-supervised learning for fine renal artery segmentation," *Medical Image Analysis*, vol. 63, p. 101722, 2020.
- [35] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han *et al.*, "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge," *Medical Image Analysis*, vol. 67, p. 101821, 2021.

- [36] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14–18, 2014, Proceedings, Part I* 17, 2014, pp. 520–527.
- [37] H. Roth, L. Le, S. Ari, K. Cherry, J. Hoffman, S. Wang, and R. Summers, "A new 2.5 d representation for lymph node detection in ct," *The Cancer Imaging Archive*, 2018.
- [38] X. Zhuang, "Multivariate mixture model for myocardial segmentation combining multi-source images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 2933–2946, 2018.
- [39] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of mri," *Medical Image Analysis*, vol. 31, pp. 77–87, 2016.
- [40] X. Luo and X. Zhuang, "X-metric: An n-dimensional information-theoretic framework for groupwise registration and deep combined computing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [41] X. Zhuang, L. Li, C. Payer, D. Štern, M. Urschler, M. P. Heinrich, J. Oster, C. Wang, Ö. Smedby, C. Bian *et al.*, "Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge," *Medical Image Analysis*, vol. 58, p. 101537, 2019.
- [42] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers *et al.*, "The medical segmentation decathlon," *Nature Communications*, vol. 13, no. 1, p. 4128, 2022.
- [43] K. J. Kiser, S. Ahmed, S. Stieb, A. S. Mohamed, H. Elhalawani, P. Y. Park, N. S. Doyle, B. J. Wang, A. Barman, Z. Li *et al.*, "Plethora: Pleural effusion and thoracic cavity segmentations in diseased lungs for benchmarking chest ct processing pipelines," *Medical Physics*, vol. 47, no. 11, pp. 5941–5952, 2020.
- [44] K. J. Kiser, S. Ahmed, S. M. Stieb, A. S. Mohamed, H. Elhalawani, P. Y. Park, N. S. Doyle, B. J. Wang, A. Barman, C. D. Fuller, and L. Giancardo, "Data from the thoracic volume and pleural effusion segmentations in diseased lungs for benchmarking chest ct processing pipelines (plethora)," Data set, 2020. [Online]. Available: <https://doi.org/10.7937/tcia.2020.6c7y-gq39>
- [45] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, J. A. Benson, W. Zhang, A. N. Leung *et al.*, "A radiogenomic dataset of non-small cell lung cancer," *Scientific Data*, vol. 5, no. 1, pp. 1–9, 2018.
- [46] B. Rister, D. Yi, K. Shivakumar, T. Nobashi, and D. L. Rubin, "Ct-org, a new dataset for multiple organ segmentation in computed tomography," *Scientific Data*, vol. 7, no. 1, p. 381, 2020.
- [47] Z. Lambert, C. Petitjean, B. Dubray, and S. Kuan, "Segthor: Segmentation of thoracic organs at risk in ct images," in *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2020, pp. 1–6.
- [48] "Structseg2019 grand challenge dataset," Website, retrieved from <https://structseg2019.grand-challenge.org/>.
- [49] J. Wasserthal, M. Meyer, H.-C. Breit, J. Cyriac, S. Yang, and M. Segeroth, "Totalsegmentator: robust segmentation of 104 anatomical structures in ct images," *arXiv preprint arXiv:2208.05868*, 2022.
- [50] X. Luo, W. Liao, J. Xiao, J. Chen, T. Song, X. Zhang, K. Li, D. N. Metaxas, G. Wang, and S. Zhang, "Word: A large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from ct image," *Medical Image Analysis*, vol. 82, p. 102642, 2022.
- [51] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Çetin, K. Lekadir, O. Camara, M. A. G. Ballester *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.
- [52] S.-L. Liew, B. P. Lo, M. R. Donnelly, A. Zavaliangos-Petropulu, J. N. Jeong, G. Barisano, A. Hutton, J. P. Simon, J. M. Juliano, A. Suri *et al.*, "A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms," *Scientific Data*, vol. 9, no. 1, p. 320, 2022.
- [53] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, and Q. Feng, "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PloS One*, vol. 10, no. 10, p. e0140381, 2015.
- [54] J. Cheng, W. Yang, M. Huang, W. Huang, J. Jiang, Y. Zhou, R. Yang, J. Zhao, Y. Feng, Q. Feng *et al.*, "Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation," *PloS One*, vol. 11, no. 6, p. e0157112, 2016.
- [55] Y. Wang, W. Duggar, D. Caballero, T. Vengaloor Thomas, N. Adari, E. Mundra, and H. Wang, "Brain tumor recurrence prediction after gamma knife radiotherapy from mri and related dicom-rt: An open annotated dataset and baseline algorithm (brain-tr-gammaknife) [dataset]," *The Cancer Imaging Archive*, 2023.
- [56] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.
- [57] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint arXiv:1811.02629*, 2018.
- [58] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycski, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, p. 170117, 2017.
- [59] B. Spyridon, A. Hamed, S. Aristeidis, B. Michel, R. Martin, K. Justin, F. John, F. Keyvan, and D. Christos, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection," *The Cancer Imaging Archive*, 2017.
- [60] ———, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection," *The Cancer Imaging Archive*, 2017.
- [61] S. R. Bowen, W. T. Yuh, D. S. Hippe, W. Wu, S. C. Partridge, S. Elias, G. Jia, Z. Huang, G. A. Sandison, D. Nelson *et al.*, "Tumor radiomic heterogeneity: multiparametric functional imaging to characterize variability and predict response following cervical cancer radiation therapy," *Journal of Magnetic Resonance Imaging*, vol. 47, no. 5, pp. 1388–1396, 2018.
- [62] A. E. Kavur, N. S. Gezer, M. Barış, S. Aslan, P.-H. Conze, V. Groza, D. D. Pham, S. Chatterjee, P. Ernst, S. Özkan, B. Baydar, D. Lachinov, S. Han, J. Pauli, F. Isensee, M. Perkoniig, R. Sathish, R. Rajan, D. Sheet, G. Dovletov, O. Speck, A. Nürnberg, K. H. Maier-Hein, G. Bozdagi Akar, G. Ünal, O. Dicle, and M. A. Selver, "CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation," *Medical Image Analysis*, vol. 69, p. 101950, 2021. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1361841520303145>
- [63] R. Dorent, A. Kujawa, M. Ivory, S. Bakas, N. Rieke, S. Joutard, B. Glockner, J. Cardoso, M. Modat, K. Batmanghelich *et al.*, "Crossmoda 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwannoma and cochlea segmentation," *Medical Image Analysis*, vol. 83, p. 102628, 2023.
- [64] G. Podobnik, P. Strojan, P. Peterlin, B. Ibragimov, and T. Vrtovec, "Han-seg: The head and neck organ-at-risk ct & mr segmentation dataset," *Medical Physics*, 2023.
- [65] M. R. Hernandez Petzsche, E. de la Rosa, U. Hanning, R. Wiest, W. Valenzuela, M. Reyes, M. Meyer, S.-L. Liew, F. Kofler, I. Ezhov *et al.*, "Isles 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset," *Scientific Data*, vol. 9, no. 1, p. 762, 2022.
- [66] R. Trigui, J. Mitérán, P. M. Walker, L. Sellami, and A. B. Hamida, "Automatic classification and localization of prostate cancer using multi-parametric mri/mrs," *Biomedical Signal Processing and Control*, vol. 31, pp. 189–198, 2017.
- [67] A. Vassantachart, Y. Cao, M. Gribble, S. Guzman, J. C. Ye, K. Hurth, A. Mathew, G. Zada, Z. Fan, E. L. Chang *et al.*, "Automatic differentiation of grade i and ii meningiomas on magnetic resonance image using an asymmetric convolutional neural network," *Scientific Reports*, vol. 12, no. 1, p. 3806, 2022.
- [68] V. M. Campello, P. Gkontra, C. Izquierdo, C. Martin-Isla, A. Sojoudi, P. M. Full, K. Maier-Hein, Y. Zhang, Z. He, J. Ma *et al.*, "Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3543–3554, 2021.

- [69] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019.
- [70] N. Bloch, A. Madabhushi, H. Huisman, J. Freymann, J. Kirby, M. Grauer, A. Enquobahrie, C. Jaffe, L. Clarke, and K. Farahani, "Nci-isbi 2013 challenge: Automated segmentation of prostate structures," *The Cancer Imaging Archive*, 2015.
- [71] A. Saha, M. Hosseinzadeh, and H. Huisman, "End-to-end prostate cancer detection in bpmri via 3d cnns: effects of attention mechanisms, clinical priori and decoupled false positive reduction," *Medical Image Analysis*, vol. 73, p. 102155, 2021.
- [72] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury *et al.*, "The parkinson progression marker initiative (ppmi)," *Progress in Neurobiology*, vol. 95, no. 4, pp. 629–635, 2011.
- [73] G. Litjens, R. Toth, W. Van De Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang *et al.*, "Evaluation of prostate segmentation algorithms for mri: the promise12 challenge," *Medical Image Analysis*, vol. 18, no. 2, pp. 359–373, 2014.
- [74] A. Fedorov, M. Schwier, D. Clunie, C. Herz, S. Pieper, R. Kikinis, C. Tempany, and F. Fennessy, "An annotated test-retest collection of prostate multiparametric mri," *Scientific Data*, vol. 5, no. 1, pp. 1–13, 2018.
- [75] A. S. Becker, K. Chaitanya, K. Schawkat, U. J. Muehlematter, A. M. Hötker, E. Konukoglu, and O. F. Donati, "Variability of manual segmentation of the prostate in axial t2-weighted mri: A multi-reader study," *European Journal of Radiology*, vol. 121, p. 108716, 2019.
- [76] D. Zukić, A. Vlasák, J. Egger, D. Hořínek, C. Nimsky, and A. Kolb, "Robust detection and segmentation for diagnosis of vertebral diseases using routine mr images," in *Computer Graphics Forum*, vol. 33, no. 6, 2014, pp. 190–204.
- [77] H. J. Kuijf, J. M. Biesbroek, J. De Bresser, R. Heinen, S. Andermatt, M. Bento, M. Berseth, M. Belyaev, M. J. Cardoso, A. Casamitjana *et al.*, "Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge," *IEEE Transactions on Medical Imaging*, vol. 38, no. 11, pp. 2556–2568, 2019.
- [78] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani *et al.*, "Automatic tuberculosis screening using chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 233–245, 2013.
- [79] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 577–590, 2013.
- [80] vbookshelf, "Chest x-ray images with pneumothorax masks," <https://www.kaggle.com/datasets/vbookshelf/pneumothorax-chest-xray-images-and-masks>, 2020.
- [81] S. for Imaging Informatics in Medicine (SIIM), "Siim-acr pneumothorax segmentation," <https://www.kaggle.com/competitions/siim-acr-pneumothorax-segmentation/overview>, 2019.
- [82] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaiher, M. S. Khan *et al.*, "Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images," *Computers in Biology and Medicine*, vol. 132, p. 104319, 2021.
- [83] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al Emadi *et al.*, "Can ai help in screening viral and covid-19 pneumonia?" *Ieee Access*, vol. 8, pp. 132 665–132 676, 2020.
- [84] A. M. Tahir, M. E. Chowdhury, A. Khandakar, T. Rahman, Y. Qiblawey, U. Khurshid, S. Kiranyaz, N. Ibtehaz, M. S. Rahman, S. Al-Maadeed *et al.*, "Covid-19 infection localization and severity grading from chest x-ray images," *Computers in Biology and Medicine*, vol. 139, p. 105002, 2021.
- [85] A. Degerli, M. Ahishali, M. Yamac, S. Kiranyaz, M. E. Chowdhury, K. Hameed, T. Hamid, R. Mazhar, and M. Gabbouj, "Covid-19 infection map generation and detection from chest x-ray images," *Health Information Science and Systems*, vol. 9, no. 1, p. 15, 2021.
- [86] A. M. Tahir, M. E. H. Chowdhury, Y. Qiblawey, A. Khandakar, T. Rahman, S. Kiranyaz, U. Khurshid, N. Ibtehaz, S. Mahmud, and M. Ezeddin, "Covid-qu-ex dataset," 2022.
- [87] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.
- [88] ———, "Development of a digital image database for chest radiographs with and without a lung nodule," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.
- [89] B. van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006.
- [90] D. Barušauskas. Ranzcr Clip lung contours. [Online]. Available: <https://www.kaggle.com/datasets/raddar/ranzcr-clip-lung-contours>
- [91] A. Degerli, S. Kiranyaz, M. E. Chowdhury, and M. Gabbouj, "Osegnnet: Operational segmentation network for covid-19 detection using chest x-ray images," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 2306–2310.
- [92] R. Khaled, M. Helal, O. Alfarghaly, O. Mokhtar, A. Elkorany, H. El Kassas, and A. Fahmy, "Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research," *Scientific Data*, vol. 9, no. 1, p. 122, 2022.
- [93] ———, "Categorized digital database for low energy and subtracted contrast enhanced spectral mammography images [dataset]," 2021.
- [94] Z. Ahmed, S. Q. Panhwar, A. Baqai, F. A. Umran, M. Ahmed, and A. Khan, "Deep learning based automated detection of intraretinal cystoid fluid," *International Journal of Imaging Systems and Technology*, vol. 32, no. 3, pp. 902–917, 2022.
- [95] Z. Ahmed, M. Ahmed, A. Baqai, and F. A. Umran, "Intraretinal cystoid fluid," 2022.
- [96] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomedical Optics Express*, vol. 6, no. 4, pp. 1172–1194, 2015.
- [97] S. Vitale, J. I. Orlando, E. Iarussi, and I. Larribide, "Improving realism in patient-specific abdominal ultrasound simulation using cyclegans," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 2, pp. 183–192, 2020.
- [98] W. Al-Dhabayani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, p. 104863, 2020.
- [99] S. Leclerc, E. Smistad, J. Pedrosa, A. Østvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P.-M. Jodoin, T. Grenier *et al.*, "Deep learning for segmentation using an open large-scale dataset in 2d echocardiography," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2198–2210, 2019.
- [100] Y. Song, J. Zheng, L. Lei, Z. Ni, B. Zhao, and Y. Hu, "Ct2us: Cross-modal transfer learning for kidney segmentation in ultrasound images with synthesized data," *Ultrasoundics*, vol. 122, p. 106706, 2022.
- [101] Y. Lu, M. Zhou, D. Zhi, M. Zhou, X. Jiang, R. Qiu, Z. Ou, H. Wang, D. Qiu, M. Zhong *et al.*, "The jnu-ifm dataset for segmenting pubic symphysis-fetal head," *Data in Brief*, vol. 41, p. 107904, 2022.
- [102] T. L. van den Heuvel, D. de Bruijn, C. L. de Korte, and B. v. Ginneken, "Automated measurement of fetal head circumference using 2d ultrasound images," *PloS One*, vol. 13, no. 8, p. e0200412, 2018.
- [103] J. Zhou, X. Jia, Y. Dong, D. Ni, A. Noble, R. Huang, T. Tan, X. Tao, R. Li, and M. The Van, "Thyroid nodule segmentation and classification," <https://tn-scui2020.grand-challenge.org/Home/>, 2020.
- [104] Kaggle, "Ultrasound nerve segmentation," <https://www.kaggle.com/competitions/ultrasound-nerve-segmentation/data>, 2016.
- [105] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, pp. 1–9, 2018.

- [106] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 168–172.
- [107] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti *et al.*, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1902.03368*, 2019.
- [108] University of Waterloo, "Skin Cancer Detection," <https://uwaterloo.ca/vision-image-processing-lab/research-demos/skin-cancer-detection>.
- [109] P. Ngoc Lan, N. S. An, D. V. Hang, D. V. Long, T. Q. Trung, N. T. Thuy, and D. V. Sang, "Neounet: Towards accurate colon polyp segmentation and neoplasm detection," in *Advances in Visual Computing: 16th International Symposium, ISVC 2021, Virtual Event, October 4–6, 2021, Proceedings, Part II*, 2021, pp. 15–28.
- [110] W.-Y. Hong, C.-L. Kao, Y.-H. Kuo, J.-R. Wang, W.-L. Chang, and C.-S. Shih, "Cholecseg8k: a semantic segmentation dataset for laparoscopic cholecystectomy based on cholec80," *arXiv preprint arXiv:2012.12453*, 2020.
- [111] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. De Mathelin, and N. Padoy, "Endonet: a deep architecture for recognition tasks on laparoscopic videos," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 86–97, 2016.
- [112] D. Jha, P. H. Smedsrød, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-seg: A segmented polyp dataset," in *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II* 26, 2020, pp. 451–462.
- [113] S. Maqbool, A. Riaz, H. Sajid, and O. Hasan, "m2caiseg: Semantic segmentation of laparoscopic images using convolutional neural networks," *arXiv preprint arXiv:2008.10134*, 2020.
- [114] S. Ali, D. Jha, N. Ghatwary, S. Realdon, R. Cannizzaro, O. E. Salem, D. Lamarque, C. Daul, M. A. Riegler, K. V. Anonsen *et al.*, "Polypgen: A multi-center polyp detection and segmentation dataset for generalisability assessment," *arXiv preprint arXiv:2106.04463*, 2021.
- [115] S. Ali, N. Ghatwary, D. Jha, E. Isik-Polat, G. Polat, C. Yang, W. Li, A. Galdran, M.-A. G. Ballester, V. Thambawita *et al.*, "Assessing generalisability of deep learning-based polyp detection and segmentation methods through a computer vision challenge," *arXiv preprint arXiv:2202.12031*, 2022.
- [116] S. Ali, M. Dmitrieva, N. Ghatwary, S. Bano, G. Polat, A. Temizel, A. Krenzer, A. Hekalo, Y. B. Guo, B. Matuszewski *et al.*, "Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy," *Medical Image Analysis*, vol. 70, p. 102002, 2021.
- [117] L. C. Garcia-Peraza-Herrera, L. Fidon, C. D'Etorre, D. Stoyanov, T. Vercauteren, and S. Ourselin, "Image compositing for segmentation of surgical tools without manual annotations," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1450–1460, 2021.
- [118] J. Yoon, S. Hong, S. Hong, J. Lee, S. Shin, B. Park, N. Sung, H. Yu, S. Kim, S. Park *et al.*, "Surgical scene segmentation using semantic image synthesis with a virtual surgery environment," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VII*, 2022, pp. 551–561.
- [119] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, 2018.
- [120] O. Kovalyk, J. Morales-Sánchez, R. Verdú-Monedero, I. Sellés-Navarro, A. Palazón-Cabanes, and J.-L. Sancho-Gómez, "Papila: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment," *Scientific Data*, vol. 9, no. 1, p. 291, 2022.
- [121] F. Li, D. Song, H. Chen, J. Xiong, X. Li, H. Zhong, G. Tang, S. Fan, D. S. Lam, W. Pan *et al.*, "Development and clinical deployment of a smartphone-based visual field deep learning system for glaucoma detection," *NPJ Digital Medicine*, vol. 3, no. 1, p. 123, 2020.
- [122] J. I. Orlando, H. Fu, J. B. Breda, K. Van Keer, D. R. Bathula, A. Diaz-Pinto, R. Fang, P.-A. Heng, J. Kim, J. Lee *et al.*, "Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Medical Image Analysis*, vol. 59, p. 101570, 2020.
- [123] K. Sirinukunwattana, D. R. Snead, and N. M. Rajpoot, "A stochastic polygons model for glandular structures in colon histology images," *IEEE Transactions on Medical Imaging*, vol. 34, no. 11, pp. 2366–2378, 2015.
- [124] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez *et al.*, "Gland segmentation in colon histology images: The glas challenge contest," *Medical Image Analysis*, vol. 35, pp. 489–502, 2017.
- [125] HuBMAP + HPA, "HuBMAP + HPA - Hacking the Human Body," <https://www.kaggle.com/competitions/hubmap-organ-segmentation>, 2022.
- [126] InnovationDigi, "HuBMAP - Hacking the Kidney: Identify glomeruli in human kidney tissue images," <https://www.kaggle.com/competitions/hubmap-kidney-segmentation/data?select=test>, 2021.