

## 1 A Dataset Details

2 In this section, we provide the detailed datasets used in MedSG-Bench, including the name of  
3 the dataset, the modality, the dimension of data, and the accessible links. As shown in Table 1,  
4 MedSG-Bench is constructed from 76 datasets across 10 medical image modalities.

Table 1: Detailed datasets information in MedSG-Bench.

Dataset	Modality	Dim	Accessible links
4C2021	CT	3D	<a href="https://aistudio.baidu.com/datasetdetail/89548">https://aistudio.baidu.com/datasetdetail/89548</a>
AbdomenCT1K	CT	3D	<a href="https://github.com/JunMa11/AbdomenCT-1K">https://github.com/JunMa11/AbdomenCT-1K</a>
ACDC	MRI	3D	<a href="https://humanheart-project.creatis.insa-lyon.fr/database/">https://humanheart-project.creatis.insa-lyon.fr/database/</a>
AMOS22	CT, MRI	3D	<a href="https://amos22.grand-challenge.org/">https://amos22.grand-challenge.org/</a>
ATM22	CT	3D	<a href="https://atm22.grand-challenge.org/">https://atm22.grand-challenge.org/</a>
Atria Segmentation	MRI	3D	<a href="https://www.cardiacatlas.org/atriaseg2018-challenge/atria-seg-data/">https://www.cardiacatlas.org/atriaseg2018-challenge/atria-seg-data/</a>
AutoLaparo	Colonoscopy	2D	<a href="https://autolaparo.github.io/">https://autolaparo.github.io/</a>
BAGLS	Endoscopy	2D	<a href="https://www.kaggle.com/datasets/gomezp/benchmark-for-automatic-glottis-segmentation">https://www.kaggle.com/datasets/gomezp/benchmark-for-automatic-glottis-segmentation</a>
BraimMRI	MRI	3D	<a href="https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset">https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset</a>
BrainPTM	MRI	3D	<a href="https://brainptm-2021.grand-challenge.org/">https://brainptm-2021.grand-challenge.org/</a>
BraTS2020	MRI	3D	<a href="https://service.tib.eu/ldmservice/dataset/brats2020">https://service.tib.eu/ldmservice/dataset/brats2020</a>
BUSI	US	2D	<a href="https://scholar.cu.edu.eg/?q=afahmy/pages/dataset">https://scholar.cu.edu.eg/?q=afahmy/pages/dataset</a>
CAD-PE	CT	3D	<a href="https://ieee-dataport.org/open-access/cad-pe">https://ieee-dataport.org/open-access/cad-pe</a>
CAMUS	US	2D	<a href="https://www.creatis.insa-lyon.fr/Challenge/camus/">https://www.creatis.insa-lyon.fr/Challenge/camus/</a>
Cause07	MRI	3D	<a href="https://cause07.grand-challenge.org/">https://cause07.grand-challenge.org/</a>
CBCT3D	CBCT	3D	<a href="https://toothfairy.grand-challenge.org/">https://toothfairy.grand-challenge.org/</a>
Chestimage	X-Ray	2D	<a href="https://tianchi.aliyun.com/dataset/83075">https://tianchi.aliyun.com/dataset/83075</a>
CMRxMotions	MRI	3D	<a href="https://www.synapse.org/Synapse:syn28503327/">https://www.synapse.org/Synapse:syn28503327/</a>
COVID-19	CT	3D	<a href="https://medicalsegmentation.com/covid19/">https://medicalsegmentation.com/covid19/</a>
COVID19CTscans	CT	3D	<a href="https://zenodo.org/records/3757476">https://zenodo.org/records/3757476</a>
COVID-19-20	CT	3D	<a href="https://covid-segmentation.grand-challenge.org/">https://covid-segmentation.grand-challenge.org/</a>
Covid19cxr	X-ray	2D	<a href="https://github.com/ieee8023/covid-chestxray-dataset">https://github.com/ieee8023/covid-chestxray-dataset</a>
Cranium	CT	3D	<a href="https://tianchi.aliyun.com/dataset/82967">https://tianchi.aliyun.com/dataset/82967</a>
CT-ORG	CT	3D	<a href="https://www.cancerimagingarchive.net/collection/ct-org/">https://www.cancerimagingarchive.net/collection/ct-org/</a>
CTSpine1K	CT	3D	<a href="https://github.com/MIRACLE-Center/CTSpine1K">https://github.com/MIRACLE-Center/CTSpine1K</a>
CVC-ClinicDB	Colonoscopy	2D	<a href="https://polyp.grand-challenge.org/CVCClinicDB/">https://polyp.grand-challenge.org/CVCClinicDB/</a>
DRISHTI-GS	Fundus	2D	<a href="https://www.kaggle.com/datasets/lokeshaipureddi/drishtigs-retina-dataset-for-onh-segmentation">https://www.kaggle.com/datasets/lokeshaipureddi/drishtigs-retina-dataset-for-onh-segmentation</a>
EMIDEC	MRI	3D	<a href="https://emidec.com/dataset">https://emidec.com/dataset</a>
EndoTect2020	Colonoscopy	2D	<a href="https://osf.io/mh9sj/">https://osf.io/mh9sj/</a>
EndoVis15	Colonoscopy	2D	<a href="https://endovis.grand-challenge.org/">https://endovis.grand-challenge.org/</a>
EndoVis2017	Colonoscopy	2D	<a href="https://endovissub2017-roboticinstrumentsegmentation.grand-challenge.org/">https://endovissub2017-roboticinstrumentsegmentation.grand-challenge.org/</a>
GAMMA	Fundus	2D	<a href="https://gamma.grand-challenge.org/Home/">https://gamma.grand-challenge.org/Home/</a>
HaN-Seg	CT, MRI	3D	<a href="https://zenodo.org/records/7442914">https://zenodo.org/records/7442914</a>
Hvsmr2016	MRI	3D	<a href="http://segchd.csail.mit.edu/data.html">http://segchd.csail.mit.edu/data.html</a>
I2CVB	MRI	3D	<a href="https://i2cvb.github.io/">https://i2cvb.github.io/</a>
InSTANCE2022	CT	3D	<a href="https://instance.grand-challenge.org/">https://instance.grand-challenge.org/</a>
iseg2017	MRI	3D	<a href="https://iseg2017.web.unc.edu/download/">https://iseg2017.web.unc.edu/download/</a>
ISIC2018	Dermoscopy	2D	<a href="https://challenge.isic-archive.com/data/#2018">https://challenge.isic-archive.com/data/#2018</a>
ISLES-ATLAS	MRI	3D	<a href="https://atlas.grand-challenge.org/">https://atlas.grand-challenge.org/</a>
ISLES-MM	MRI	3D	<a href="https://isles22.grand-challenge.org/">https://isles22.grand-challenge.org/</a>
JSRT	X-ray	2D	<a href="http://imgcom.jsrt.or.jp/minijsrtdb/">http://imgcom.jsrt.or.jp/minijsrtdb/</a>
KvasirInstrument	Colonoscopy	2D	<a href="https://datasets.simula.no/kvasir-instrument/">https://datasets.simula.no/kvasir-instrument/</a>
LMSLS	MRI	3D	<a href="https://smart-stats-tools.org/lesion-challenge-2015">https://smart-stats-tools.org/lesion-challenge-2015</a>

LUNA16	CT	3D	<a href="https://luna16.grand-challenge.org/Download/">https://luna16.grand-challenge.org/Download/</a>
MMWHS	CT, MRI	3D	<a href="https://www.ub.edu/mnms/">https://www.ub.edu/mnms/</a>
MRSpineSeg	MRI	3D	<a href="https://mosmed.ai/datasets/covid19_1110">https://mosmed.ai/datasets/covid19_1110</a>
MSD02	MRI	3D	<a href="http://medicaldecathlon.com/">http://medicaldecathlon.com/</a>
MSD04	MRI	3D	<a href="http://medicaldecathlon.com/">http://medicaldecathlon.com/</a>
MSD05	MRI	3D	<a href="http://medicaldecathlon.com/">http://medicaldecathlon.com/</a>
MyoPS2020	MRI	3D	<a href="https://zmiclab.github.io/zxh/0/myops20/">https://zmiclab.github.io/zxh/0/myops20/</a>
NCI-ISBI2013	MRI	3D	<a href="https://www.cancerimagingarchive.net/analysis-result/isbi-mr-prostate-2013/">https://www.cancerimagingarchive.net/analysis-result/isbi-mr-prostate-2013/</a>
PadChest	X-ray	2D	<a href="https://bimcv.cipf.es/bimcv-projects/padchest/">https://bimcv.cipf.es/bimcv-projects/padchest/</a>
PALM	Fundus	2D	<a href="https://ieee-dataport.org/documents/palm-pathologic-myopia-challenge">https://ieee-dataport.org/documents/palm-pathologic-myopia-challenge</a>
Parse2022	CT	3D	<a href="https://parse2022.grand-challenge.org/Dataset/">https://parse2022.grand-challenge.org/Dataset/</a>
PCXA	X-ray	2D	<a href="https://lhncbc.nlm.nih.gov/LHC-downloads/downloads.html">https://lhncbc.nlm.nih.gov/LHC-downloads/downloads.html</a>
PDDCA	CT	3D	<a href="https://www.imagenglab.com/newsite/pddca/">https://www.imagenglab.com/newsite/pddca/</a>
Pelvic1K	CT	3D	<a href="https://zenodo.org/record/4588403">https://zenodo.org/record/4588403</a>
Promise09	MRI	3D	<a href="https://www.na-mic.org/wiki/Training_Data_Prostate_Segmentation_Challenge_MICCAI09">https://www.na-mic.org/wiki/Training_Data_Prostate_Segmentation_Challenge_MICCAI09</a>
PROMISE12	MRI	3D	<a href="https://zenodo.org/records/8026660">https://zenodo.org/records/8026660</a>
QaTa-COV19	X-ray	2D	<a href="https://www.kaggle.com/datasets/ayseenderli/qatacov19-dataset">https://www.kaggle.com/datasets/ayseenderli/qatacov19-dataset</a>
QUBIQ2020	CT	2D	<a href="https://qubiq.grand-challenge.org/">https://qubiq.grand-challenge.org/</a>
REFUGE	Fundus	2D	<a href="https://refuge.grand-challenge.org/">https://refuge.grand-challenge.org/</a>
RIGA+	Fundus	2D	<a href="https://zenodo.org/records/6325549">https://zenodo.org/records/6325549</a>
RIM_ONE	Fundus	2D	<a href="https://github.com/miag-ull/rim-one-dl">https://github.com/miag-ull/rim-one-dl</a>
SegRap2023	CT	2D	<a href="https://segrap2023.grand-challenge.org/dataset/">https://segrap2023.grand-challenge.org/dataset/</a>
SegTHOR	CT	3D	<a href="https://competitions.codalab.org/competitions/21145">https://competitions.codalab.org/competitions/21145</a>
SIIM-ACR	X-ray	2D	<a href="https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation">https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation</a>
SKI10	MRI	3D	<a href="https://ski10.grand-challenge.org/">https://ski10.grand-challenge.org/</a>
SLAWT	MRI	3D	<a href="http://stacom.cardiacatlas.org/">http://stacom.cardiacatlas.org/</a>
TBAD	CTA	3D	<a href="https://www.kaggle.com/datasets/xiaoweixumedaicalai/imagetbad">https://www.kaggle.com/datasets/xiaoweixumedaicalai/imagetbad</a>
TN-SCUI	US	2D	<a href="https://tn-scui2020.grand-challenge.org/">https://tn-scui2020.grand-challenge.org/</a>
VESSEL12	CT	3D	<a href="https://vessel12.grand-challenge.org/">https://vessel12.grand-challenge.org/</a>
VINDR-Mammo	X-ray	2D	<a href="https://www.physionet.org/content/vindr-mammo/1.0.0/">https://www.physionet.org/content/vindr-mammo/1.0.0/</a>
Verse19	CT	3D	<a href="https://github.com/anjany/verse">https://github.com/anjany/verse</a>
WMH	MRI	3D	<a href="https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/AECRS">https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/AECRS</a>
WORD	CT	3D	<a href="https://github.com/HiLab-git/WORD">https://github.com/HiLab-git/WORD</a>

## 5 B Template for Instruction Data Generation

### Template for Instruction Data Generation

#### Task1

"Please examine these two images and provide the coordinates of the area where they differ."

"Compare both images closely and share the coordinates of the discrepancy."

"Look at these two images and tell me the coordinates of the difference between them."

"Carefully analyze these images and provide the coordinates of their difference."

"Examine the two images and give me the coordinates of the region where they differ."

"Can you find the differences between these two images and give me the coordinates?"

"Please inspect these two images and indicate the coordinates of their difference."

"Compare the two images and identify the coordinates of the difference."

"Look closely at the two images and provide the coordinates where they differ."

"Analyze both images and provide the coordinates of the difference between them."

#### Task2

"Compare these two images carefully and give me the coordinates of their real difference in the second image. Find it and locate it in the second image."

"Please examine both images and identify the real difference that appears in the second one. Provide the coordinates of that difference."

"Carefully analyze the two images. What is the actual visual change in the second image? Mark its coordinates precisely."

"Spot the true difference in the second image when compared with the first. Return the bounding box of that change."

"Look at the two images side by side. What is the meaningful change introduced in the second image? Output its location."

"Your task is to detect the actual difference in the second image compared to the first and report its position in coordinates."

"Inspect the two images and tell me where the real change is in the second one. Output the coordinates of the difference."

"Between the two images, find the true variation that exists in the second image. Return its location in bounding box format."

"Compare the pair of images. Where is the real and only difference in the second image? Provide the coordinates."

"Analyze the difference between these images. Identify and locate the actual modified region in the second image only."

#### Task3

"The object marked with a red bounding box in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) is shared by these two images. Locate and identify it in the num image."

"In the first image, the object highlighted with a red bounding box (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) is common to both images. Please recognize and locate it in the num image."

"The object outlined by a red bounding box in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) appears in both images. Can you identify and find its position in the num image?"

"The object with a red bounding box in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) is shared between these two images. Locate and recognize it in the num image."

"The object marked in red in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) is common across both images. Find and identify it in the num image."

"The object highlighted by the red box in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) is shared with the second image. Locate it in the num image and provide its position."

"Both images contain a common object marked with a red bounding box in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>). Find and identify this object in the num image."

"In the first image, the object marked by the red bounding box (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) appears in both. Can you locate it in the num image?"

"The object in the first image, marked by a red bounding box (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>), is also in the second image. Identify and locate it in the num image."

"In the first image, the object enclosed by the red bounding box (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) is the same as in the second image. Locate it in the num image and identify its position."

#### Template for Instruction Data Generation

##### Task4

"In the first image, a red bounding box marks a specific object (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>). Your task is to identify and localize the same object in the num image."

"The object enclosed in red in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) also appears in the num image. Detect and locate it accordingly."

"Focus on the object highlighted by the red box in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>). Find and mark this same object in the num image."

"Observe the red-boxed object in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>). Identify where it appears in the num image."

"The first image contains an object inside a red bounding box (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>). Detect this same object in the num image."

"An object is annotated with a red box in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>). Determine where the same object appears in the num image."

"Use the red-bounded object in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>) as a reference. Identify its location in the num image."

"Locate in the num image the object that corresponds to the red-marked region in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>)."

"The first image includes an object shown with a red bounding box (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>). Recognize and localize this same object in the num image."

"Refer to the red-outlined region in the first image (<|box\_start|> (x\_min, y\_min), (x\_max, y\_max) <|box\_end|>). Locate the corresponding object in the num image."

##### Task5

"Given image-1 and image-2, identify and localize the object from image-1 within image-2."

"Based on the object shown in image-1, determine its corresponding location in image-2."

"Observe the object in image-1. Where does it appear in image-2? Mark the location."

"Find the region in image-2 that corresponds to the object highlighted in image-1."

"Refer to image-1 and locate the same object in image-2."

"Your task is to recognize the object from image-1 and indicate where it is in image-2."

"Using image-1 as a reference, identify the location of the same object in image-2."

"Locate the counterpart of the object shown in image-1 within image-2."

"Match the object in image-1 to its corresponding region in image-2 and provide its location."

"Analyze the object in image-1 and find its equivalent presence in image-2 by marking its location."

##### Task6

"You are given a source image and several cropped regions. Identify where the num region belongs in the source image."

"Observe the original image and its cropped parts. Locate the num region in the source image."

"Given one complete image and multiple region crops, find where the num one fits in the original image."

"You are shown a source image and some regional cutouts. Point out where the num region comes from."

"Refer to the original image and determine the location of the num region shown afterward."

"Analyze the full image and match the num region image to its location within it."

"Based on the source image, indicate where the num region patch belongs."

"Here is a source image followed by cropped regions. Find the position of the num region in the source."

"You are given a full image and several region patches. Locate the num patch within the source image."

#### Template for Instruction Data Generation

##### **Task7**

"You are given total images. Based on the red bounding box in the first image, locate the corresponding region in the num image that shares a similar function or meaning."

"Among the total provided images, examine the red-highlighted area in the first image and identify the region in the num image that matches it semantically or functionally."

"You are given total images. Consider the red-marked region in the first image. In the num image, find the area that best aligns with it in terms of purpose or meaning."

"From the total images below, determine which region in the num image corresponds to the red-boxed area in the first image."

"You are given total images. Study the red region in the first image. Then, in the num image, identify the location that serves a similar role or conveys a similar idea."

"You are given total images. Take a close look at the red-bounded area in the first image. Locate the corresponding region in the num image that reflects the same concept."

"You are given total images. Focus on the red box in the first image. Your task is to find the equivalent region in the num image that shares its function or meaning."

"You are given total images. Analyze the highlighted region in the first image. In the num image, point out the area that represents the same functional or semantic content."

"Given total images, compare the red-boxed area in the first image with the num image and find the corresponding part."

"You are given total images. Observe the first image where a red region is marked. Identify the most similar region in the num image in terms of functionality or semantics."

##### **Task8**

"Identify the bounding box of the region described by the following expression: <object\_ref\_start> object name <object\_ref\_end>."

"Locate the region corresponding to the following structure and provide its bounding box:<object\_ref\_start> object name <object\_ref\_end>."

"What is the bounding box for the region denoted by <object\_ref\_start> object name <object\_ref\_end>?"

"Provide the bounding box for the following entity mentioned in the image: <object\_ref\_start> object name <object\_ref\_end>."

"Identify and annotate the bounding box of <object\_ref\_start> object name <object\_ref\_end>."

"Indicate the bounding box of the area that corresponds to <object\_ref\_start> object name <object\_ref\_end>."

"Determine the coordinates of the bounding box for the target structure: <object\_ref\_start> object name <object\_ref\_end>."

"What is the bounding box for the region denoted by <object\_ref\_start> object name1 <object\_ref\_end> and <object\_ref\_start> object name2 <object\_ref\_end>?"

## 6 C Data statistics of MedSG-188K

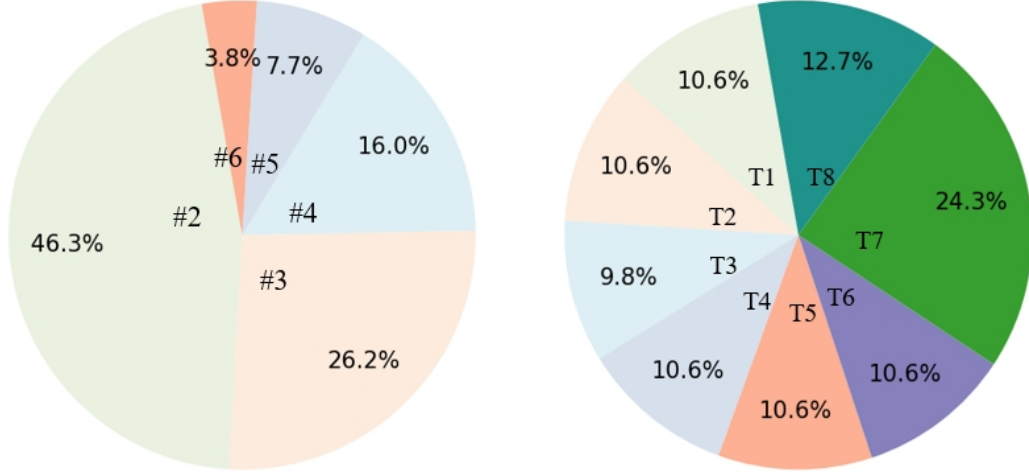


Figure 1: Proportions of image sequence length (**left**), data distribution across tasks (**right**) in MedSG-188K.

## 7 D Evaluation Metric

We evaluate model performance using two standard metrics: Intersection over Union (IoU) and Accuracy at IoU threshold 0.5 (Acc@0.5). These metrics are widely adopted in visual grounding to measure localization quality. IoU quantifies the overlap between the predicted bounding box  $B_{\text{pred}}$  and the ground-truth bounding box  $B_{\text{gt}}$ , and is defined as:

$$\text{IoU} = \frac{\text{Area}(B_{\text{pred}} \cap B_{\text{gt}})}{\text{Area}(B_{\text{pred}} \cup B_{\text{gt}})} \quad (1)$$

Acc@0.5 measures the proportion of predictions whose IoU with the ground truth exceeds 0.5. It is defined as:

$$\text{Acc@0.5} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\text{IoU}_i \geq 0.5) \quad (2)$$

Here,  $N$  is the total number of samples, and  $\mathbb{I}(\cdot)$  is the indicator function that returns 1 if the condition is true, and 0 otherwise.

## 15 E Limitations and Future Work

While MedSG-Bench is constructed from a wide range of publicly available datasets, it does not include private real-world clinical data such as longitudinal studies, multi-timepoint diagnostics, or follow-up imaging records. This limits its ability to fully capture the temporal complexity and diagnostic continuity inherent in actual clinical workflows. In future work, we plan to collaborate with medical institutions to incorporate authentic clinical data, including patient trajectories across multiple visits and imaging sessions, to enhance the benchmark’s realism and clinical applicability.

## 22 F Potential negative societal impacts

While the proposed benchmark includes eight tasks spanning medical image sequences, the resulting performance is intended for reference purposes only. High scores achieved by MLLMs on MedSG-Bench do not necessarily indicate clinical readiness or real-world applicability. Any deployment in clinical settings requires thorough validation and oversight from qualified medical professionals to ensure safety and reliability.