Fast and Low GPU Memory Abdominal Organ Segmentation in CT: Structured description of the challenge design

Remark: This challenge have been slightly modified. All changes are highlighted in red.

CHALLENGE ORGANIZATION

Title

Use the title to convey the essential information on the challenge mission.

Fast and Low GPU Memory Abdominal Organ Segmentation in CT

Challenge acronym

Preferable, provide a short acronym of the challenge (if any).

FLARE21

Challenge abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Abdominal organ segmentation plays an important role in clinical practice, and to some extent, it seems to be a solved problem because the state-of-the-art methods have achieved inter-observer performance in several benchmark datasets. However, most of the existing abdominal datasets only contain single-center, single-phase, single-vendor, or single-disease cases, and it is unclear whether the excellent performance can be generalized on more diverse datasets. Moreover, many SOTA methods use model ensembles to boost performance, but these solutions usually have a large model size and cost extensive computational resources, which are impractical to be deployed in clinical practice.
To address these limitations, we will organize a Fast and Low GPU Memory Abdominal Organ Segmentation challenge. We present a large and diverse abdominal CT organ dataset with 500 CT scans from 11 countries, including multi-center, multi-phase, multi-vendor, and multi-disease cases. Participants are required to develop segmentation methods that can segment the liver, kidney, spleen, and pancreas simultaneously, where both accuracy and efficiency will be evaluated.

The challenge has two main features: (1) the dataset is large and diverse, and the submitted methods will be evaluated on the cases from unseen centers; (2) we not only focus on segmentation accuracy but also segmentation efficiency, which are in concordance with real clinical practice and requirements.

Challenge keywords

List the primary keywords that characterize the challenge.

Abdominal organ segmentation, low gpu memory, fast, efficient

Year

The challenge will take place in ...

2021

FURTHER INFORMATION FOR MICCAI ORGANIZERS

Workshop

If the challenge is part of a workshop, please indicate the workshop.

None

Duration

How long does the challenge take?

Half day.

Expected number of participants

Please explain the basis of your estimate (e.g. numbers from previous challenges) and/or provide a list of potential participants and indicate if they have already confirmed their willingness to contribute.

Based on the number of existing challenges (e.g., KiTS 2019), we expect there will be at least 30 participants.

Publication and future plans

Please indicate if you plan to coordinate a publication of the challenge results.

We plan to coordinate a publication of challenge results with the top five participating teams.

Space and hardware requirements

Organizers of on-site challenges must provide a fair computing environment for all participants. For instance, algorithms should run on the same computing platform provided to all.

We do not have any on-site requirements.

TASK: Abdominal Organ Segmentation in CT Images

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

The purpose of this task is to benchmark multi-organ segmentation methods in terms of both accuracy and efficiency.

Keywords

List the primary keywords that characterize the task.

Segmentation, fast, liver, kidney, spleen, pancreas

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Jun Ma (Nanjing University of Science and Technology)
Song Gu (Nanjing University of Information Science and Technology)
Yao Zhang (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences)
Xingle An, China Electronics Cloud Brain (Tianjin) Technology CO., LTD.
Cheng Zhu, Shenzhen Haichuang Medical CO. LTD.
Congcong Wang (Tianjin University of Technology)
Cheng Ge (Jiangsu University of Technology)
Liwen Zou (Nanjing University)
Qiongjie Zhu (Nanjing University)
Guoqiang Dong (Nanjing University)
Jian He (Nanjing University)
Xiaoping Yang (Nanjing University)

b) Provide information on the primary contact person.

Jun Ma (junma@njust.edu.cn)

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One time event with fixed submission deadline.

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI.

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

grand-challenge.org

c) Provide the URL for the challenge website (if any).

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic.

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

No additional data allowed.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but not eligible for awards and not listed in leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

We will provide a prize (cash and certificate) for the Top 3 teams. The exact money of cash prize is not known yet. Sponsors for the awards are currently being investigated. The details of this will be announced on our challenge homepage at a later date.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Scores and rankings of all participating teams will be announced publicly.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

Participating teams are required to submit a manuscript to describe their methods, and these papers will be publicly available to community at the challenge results page.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

Submissions will be made via Docker container.

NVIDIA would offer computing resources for MICCAI 2021 challenge organizers who would like to run a Docker-based challenge. We would like to use these computing resources. However, the details are not available currently. If NVIDIA's computing resources are not available, we will use our local hardware. It has two 16-Cores CPUs, four NVIDIA TITAN XP GPUs, and 64G RAM. Only one GPU is allowed to use during inference.

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We allow five-times submission on the validation set and only once on the testing set.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

1 May, 2021 (12:00AM GMT): Launch of challenge and release of training and validation data. Docker and short paper submission of validation set opening.
1 August, 2021 (12:00AM GMT): Docker and short paper submission of testing set opening.
1 September, 2021 (12:00AM GMT): Deadline for testing submission.
27 September 27 or 1 October: Ranking of results will be available.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

Our dataset is legal to be used in this challenge, and ethics approval is not necessary.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY NC SA.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

Evaluation code will be made available with the training data.

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

Participating teams are strongly encouraged but not required to make their code publicly available

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Grants of organizers: National Natural Science Foundation of China
(No. 11971229, No. 12090023). China's Ministry of Science and Technology (No. 2020YFA0713800)
Only the organizers will have access to the test case labels.
We are currently looking into which companies are willing to sponsor the challenge in the form of awards.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

CAD.

Task category(ies)

State the task category(ies).

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Segmentation.

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

The target cohort will be patients scanned in different centres with different vendors and having different abdominal lesions.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

The challenge cohort is composed of patients with liver tumor, kidney tumor, spleen tumor, pancreas tumor as well as healthy subjects that were scanned in 12 different clinical centers.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Computed Tomography

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

The image data will be released as compressed Nifti1Image objects stored in .nii.gz files. These files will include an "affine matrix" defining the voxel size of each image.

b) ... to the patient in general (e.g. sex, medical history).

No additional information regarding each patient will be released.

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

CT including abdomen. Occasionally the chest and/or pelvis as well.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Participants have to submit a Docker image that can segment four organs: liver, kidney, spleen, and pancreas.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Precision, Runtime, Sensitivity, Hardware requirements.

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

Our dataset is adapted from MSD (Liver, Spleen, Pancreas), NIH Pancreas, KiTS, and Nanjing University under the license permission. Thus, the acquisition devices are very diverse in our dataset.

[1] Simpson, Amber L., Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram van Ginneken, Annette Kopp-Schneider, Bennett A. Landman, Geert Litjens, and Bjoern H. Menze. 2019. “A Large Annotated Medical Image Dataset for the Development and Evaluation of Segmentation Algorithms.” ArXiv Preprint ArXiv:1902.09063.
[2] Heller, Nicholas, Niranjan Sathianathen, Arveen Kalapara, Edward Walczak, Keenan Moore, Heather Kaluzniak, Joel Rosenberg et al. "The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes." arXiv preprint arXiv:1904.00445 (2019).
[3] Holger R. Roth, Amal Farag, Evrim B. Turkbey, Le Lu, Jiamin Liu, and Ronald M. Summers. (2016). Data From Pancreas-CT. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU
[4] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of

Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057. DOI: https://doi.org/10.1007/s10278-013-9622-7

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

Similar to 21a), the acquisition details are also diverse. Specifically,
MSD Pancareas dataset is obtained with the following recon-struction and acquisition parameters: pitch/table speed 0.984–1.375/39.37–27.50mm; automatic tube current modulation range, 220–380 mA; noise index, 12.5–14;120 kVp; tube rotation speed, 0.7–0.8 ms; scan delay, 80–85 s; and axial slicesreconstructed at 2.5 mm intervals.
MSD Spleen dataset is obtained with the following criteria: 120 kVp; exposure time, 500–1100 ms; and tube current, 33–440 mA. Images were reconstructed at a section thickness varying from 2.5 to 5 mm with a standard convolutional kernel and with a reconstruction di-ameter range of 360–500 mm. Iodinated contrast material (150 mL, Omnipaque300, GE Healthcare, Chicago, IL, USA) was administered intravenously for eachCT at a rate between 1 and 4 cc/s.
MSD Liver dataset is provided with the following protocol: the CT images are obtained in an in-plane resolution of 0.5 to 1.0 mm, and slice thickness of 0.45 to 6.0 mm.
KiTS dataset is provided with the following protocol: patients that contained at least one series in late arterial contrast phase that depicts the entirety of the abdomen are included.
NIH Pancreas dataset has resolutions of 512x512 pixels with varying pixel sizes and slice thickness between 1.5 2.5 mm, acquired on Philips and Siemens MDCT scanners (120 kVp tube voltage).
Nanjing University dataset is acquired with a multidetector spiral CT scanner (Lightspeed, VCT, or Discovery HD750, GE Healthcare, US). After injecting 1.2 mL/kg body weight of contrast media (Omnipaque 350 mg I/mL, GE Healthcare, US) followed by 40 mL saline solution, the arterial, portal venous and equilibrium phases were obtained at 35s, 70s and 180s, respectively. Contrast media and saline solution were injected both at a rate of 3.0 mL/s through the elbow vein by a power injector (Medrad tellant, Indianola, PA, US). CT scan parameters were as follows: tube voltage, 120 kVp; tube current, 250 - 350 mA; collimating slice thickness, 5 mm; reconstruction slice thickness, 1.25 mm; slice interval, 1.25 mm; rotation time, 0.6s; helical pitch, 1.375; field of view, 35 - 40 cm; matrix, 512 × 512. A standard reconstruction algorithm was used.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

(1) Both MSD Pancreas and Spleen datasets are provided by Memorial Sloan Kettering Cancer Center (New York, NY, USA), which can be accessed from http://medicaldecathlon.com/;
(2) MSD Liver dataset is provided from several clinical sites, including Ludwig Maximilian University of Munich (Germany), Radboud University Medical Center ofNijmegen (The Netherlands), Polytechnique and CHUM Research Center Montreal (Canada), Tel Aviv University (Israel), Sheba Medical Center (Israel), IRCAD Institute Strasbourg (France), and Hebrew University of Jerusalem (Israel) through the Liver Tumour Segmentation (LiTS) challenge, which can be accessed from https://competitions.codalab.org/competitions/17094;
(3) KiTS dataset is provided from the University of Minnesota Health;
(4) NIH Pancreas dataset is from the National Institutes of Health Clinical Center.
(5) Nanjing University dataset is from the medical school in Nanjing University.

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

Data annotation is overseen by the challenge clinical chair Dr. Jian He. He is an experienced radiologist who specializes in abdomen CT.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

One case, in this challenge, is a single CT image associated with a multi-class segmentation mask.

b) State the total number of training, validation and test cases.

The total number of cases is 511. An approximate 70%/10%/20% train/validation/testing split will be employed resulting in 361 training cases, 50 validation cases, and 100 testing cases.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

The total number of cases was mainly based on usability. The 70/10/20 split was chosen because it is standard practice in machine learning.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

The training set will contain 361 portal venous cases from the National Institutes of Health Clinical Center (80) and Memorial Sloan Kettering Cancer Center (281).

The validation set will contain 5 portal venous cases from Memorial Sloan Kettering Cancer Center, 25 late arterial phase cases from University of Minnesota, and 20 cases with diverse phases from LiTS dataset, which was collected from seven medical centers.

The testing set will contain 5 portal venous cases from Memorial Sloan Kettering Cancer Center, 25 late arterial phase cases from University of Minnesota, 20 cases with diverse phases from LiTS testing set, and 50 cases with diverse phases from Nanjing University.

All the testing cases will be hidden to participants and Docker container will be used for submisson. Also, none of the organ labels are publicly available in the testing set.

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

In the original dataset (MSD, KiTS, NIH-Pancreas), only one organ is annotated. In our dataset, we provide the annotations of four organs. Specifically, the annotations from open-source datasets were used if available, and the absent organs were annotated with the strategy of automatic segmentation and manual correction. The single-organ models were trained to infer each case. Then, 15 junior annotators manually refined the segmentation results under the supervision of 2 board-certified radiologists. Finally, 3 senior radiologists verified and refined the annotations.

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

Annotators and radiologists were asked to use ITK-SNAP 3.6 software.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Single-organ models were trained with nnUnet.
15 junior annotators each had 1~5 years of experience.
2 board-certified radiologists and 3 senior radiologists each had more than 10 years of experience.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

To make all the labels consistent across the datasets from different medical centers, an experienced radiologist finally check and revise all the annotations.

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

All the cases have been reoriented in the same direction.

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Error is unavoidable in medical image segmentation. We will have a github repository to solicit input from participants about possible errors that they discover in the dataset. This strategy was also used in MICCAI KiTS 2019.

b) In an analogous manner, describe and quantify other relevant sources of error.

None.

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Dice Similarity Coefficient, Normalized Surface Dice, Running time, and Maximum used GPU memory. All metrics will be used to compute the ranking.

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

The four metrics are complementary. Specifically, Dice Similarity Coefficient is used to measure the region error. Normalized Surface Dice is used to measure the boundary error. Running time, and maximum used GPU memory are used to measure the inference efficiency.

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Our ranking scheme includes the following three steps:
Step 1. Compute the four segmentation metrics (average DSC and NSD of four organs, runtime, maximum used GPU memory) for each testing case;
Step 2. Rank participant for each of the 100 testing cases; Each participant will have 100*4 rankings.
Step 3. Average all these rankings and then normalize them by the number of teams.

b) Describe the method(s) used to manage submissions with missing results on test cases.

Missing results on test cases, will result in the ranks for the corresponding metrics to be set to the maximum.

c) Justify why the described ranking scheme(s) was/were used.

The main benefit of this ranking scheme is that it can aggregate the metrics with different dimensions. The similar ranking scheme was also employed in MICCAI BraTS Challenge 2017-2020.
It is also transparency and fairness to the participants.

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Ranking variability will be characterized using the bootstrap.
This analysis will be performed in Python.

b) Justify why the described statistical method(s) was/were used.

The Bootstrap is a simple nonparametric method that relies on minimal assumptions.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

ADDITIONAL POINTS

References

Please include any reference important for the challenge design, for example publications on the data, the annotation process or the chosen metrics as well as DOIs referring to data or code.

[1] Simpson, Amber L., Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram van Ginneken, Annette Kopp-Schneider, Bennett A. Landman, Geert Litjens, and Bjoern H. Menze. 2019. "A Large Annotated Medical Image Dataset for the Development and Evaluation of Segmentation Algorithms." ArXiv Preprint ArXiv:1902.09063.
[2] Heller, Nicholas, Niranjan Sathianathen, Arveen Kalapara, Edward Walczak, Keenan Moore, Heather Kaluzniak, Joel Rosenberg et al. "The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes." arXiv preprint arXiv:1904.00445 (2019).
[3] Holger R. Roth, Amal Farag, Evrim B. Turkbey, Le Lu, Jiamin Liu, and Ronald M. Summers. (2016). Data From Pancreas-CT. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU
[4] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057. DOI: https://doi.org/10.1007/s10278-013-9622-7
[5] Medical Segmentation Decathlon: http://medicaldecathlon.com/

Further comments

Further comments from the organizers.