

# Towards Efficient COVID-19 CT Annotation: A Benchmark for Lung and Infection Segmentation

- [Task 1: Learning with limited annotations](#)
- [Task 2: Learning to segment COVID-19 CT scans from non-COVID-19 CT scans](#)
- [Task 3: Learning with both COVID-19 and non-COVID-19 CT scans](#)

## Motivation

Tremendous [studies](#) show that deep learning methods have potential for providing accurate and quantitative assessment of COVID-19 infection in CT scans if hundreds of well-labeled training cases are available.

However, manual delineation of lung and infection is time-consuming and labor-intensive. Thus, we set up this benchmark to explore annotation-efficient methods for COVID-19 CT scans segmentation. In particular, we focus on learning to segment left lung, right lung and infection using

- pure but limited COVID-19 CT scans;
- existing labeled lung CT dataset from other non-COVID-19 lung diseases;
- heterogeneous datasets include both COVID-19 and non-COVID-19 CT scans.

## Datasets

Download Dataset	Description	License
<a href="#">StructSeg 2019</a>	50 lung CT scans; Annotations include left lung, right lung, spinal cord, esophagus, heart, trachea and gross target volume of lung cancer.	Hold by the <a href="#">challenge organizers</a>
<a href="#">NSCLC</a>	402 lung CT scans; Annotations include left lung, right lung and pleural effusion (78 cases).	CC BY-NC
<a href="#">MSD Lung Tumor</a>	63 lung CT scans; Annotations include lung cancer.	CC BY-SA
<a href="#">COVID-19-CT-Seg</a>	20 lung CT scans from; Annotations include left lung, right lung and infections.	CC BY-NC-SA



## Segmentation Task 1: Learning with limited annotations

This task is based on the COVID-19-CT-Seg dataset with 20 cases. Three subtasks are to segment lung, infection or both of them. For each task, 5-fold cross-validation results should be reported. It should be noted that each fold only has 4 training cases, and remained 16 cases are used for testing. In other words, this is a few-shot or zero-shot segmentation task. Dataset split file and quantitative results of U-Net baseline are presented in Task1 folder.

Subtask	Training and Testing
Lung	5-fold cross validation
Infection	4 cases (20% for training)
Lung and infection	16 cases (80% for testing)

## Segmentation Task 2: Learning to segment COVID-19 CT scans from non-COVID-19 CT scans

This task is to segment lung and infection in COVID-19 CT scans. The main difficulty is that the training set and testing set differ in data distribution. Although all the datasets are lung CT, they vary in lesion types (i.e., cancer, pleural effusion, and COVID-19), patient cohorts and imaging scanners.

It should be noted that labeled COVID-19 CT scans are not allowed to be used during training. The following table presents the details of training, validation, and testing set. Name (Num.) denotes the dataset name and the number of cases in this dataset, e.g., StructSeg Lung (40) denotes that 40 cases in StructSeg dataset are used for training.

Dataset split file and quantitative results of U-Net baseline are presented in Task2 folder.

Subtask	Training	Validation	(Unseen)Testing
Lung	StructSeg Lung (40) NSCLC Lung (322)	StructSeg Lung (10) NSCLC Lung (80)	COVID-19-CT-Seg Lung (20)
Infection	MSD Lung Tumor (51) StructSeg Gross Target (40) NSCLC Plcural Effusion (62)	MSD Lung Tumor (12) StructSeg Gross Target (10) NSCLC Plcural Effusion (16)	COVID-19-CT-Seg Infection(20)

## Segmentation Task 3: Learning with both COVID-19 and non-COVID-19 CT scans

This task is also to segment lung and infection in COVID-19 CT scans, but a limited labeled COVID-19 CT scans are allowed to be used during training. For each subtask, 5-fold cross-validation results should be reported.

Dataset split file and quantitative results of U-Net baseline will be presented in Task3 folder.

Subtask	Training	Validation	Testing
Lung	StructSeg Lung (40) NSCLC Lung (322)	COVID-19-CT-Seg Lung(4) StructSeg Lung (10) NSCLC Lung (80)	COVID-19-CT-Seg Lung(16)
Infection	MSD Lung Tumor (51) StructSeg Gross Target (40) NSCLC Plcural Effusion (62)	MSD Lung Tumor (12) StructSeg Gross Target (10) NSCLC Plcural Effusion (16)	COVID-19-CT-Seg Infection(16)

## Guidelines

- We hope these tasks can serve as a benchmark for novel annotation-efficient segmentation methods of COVID-19 CT scans. Both semi-automatic (e.g., level set, graph cut...) and fully automatic methods (e.g., CNNs...) are welcome.
- Evaluation metrics are Dice similarity coefficient (DSC) and normalized surface Dice (NSD), and the python implementations are [here](#).
- In [COVID-19-CT-Seg](#) dataset, the last 10 cases from Radiopaedia have been adjusted to lung window [-1250,250], and then normalized to [0,255], we recommend to adjust the first 10 cases from Coronacases with the same method.
- Nifty format of the NSCLC dataset can be downloaded [here](#) (pw:1qop). It should be noted that all the copyrights belong to the original dataset contributors, and please also [cite the corresponding publications](#) if you use this dataset.
- 2D/3D U-Net baselines are based on [nnU-Net](#).
- 45 pretrained 3D U-Net baseline models and corresponding segmentation results are available [here](#). [Baidu Net Disk mirror](#) (pw: t5mj)
- [Github mirror](#); [Gitee mirror](#).

## TODO

- ☒ Provide pretrained 3D U-Net models by 5.6.
- ☐ Provide pretrained 2D U-Net models by 5.31.

## Acknowledgements

We thank all the organizers of MICCAI 2018 Medical Segmentation Decathlon, MICCAI 2019 Automatic Structure Segmentation for Radiotherapy Planning Challenge, [the Coronacases Initiative](#) and [Radiopaedia](#) for the publicly available lung CT dataset. We also thank [Joseph Paul Cohen](#) for providing convenient download [link](#) of 20 COVID-19 CT scans. We also thank all the contributor of [NSCLC](#) and [COVID-19-Seg-CT](#) dataset for providing annotations of lung, pleural effusion and COVID-19 infection. We also thank the organizers of [TMI Special Issue on Annotation-Efficient Deep Learning for Medical Imaging](#) because we get lots of insights from the call for papers when designing these segmentation tasks. We also thank the contributors of these great COVID-19 related resources: [COVID19\\_imaging\\_AI\\_paper\\_list](#) and [MedSeg](#). Last but not least, we thank Chen Chen, Xin Yang, and Yao Zhang for their important feedback on this benchmark.

Including the following two citations in your work would be highly appreciated.

- Jun Ma, Yixin Wang, Xingle An, Cheng Ge, Ziqi Yu, Jianan Chen, Qiongjie Zhu, Guoqiang Dong, Jian He, Zhiqiang He, Ziwei Nie, Xiaoping Yang, "Towards Efficient COVID-19 CT Annotation: A Benchmark for Lung and Infection Segmentation," arXiv preprint [arXiv:2004.12537](#), 2020

```
@article{COVID-19-SegBenchmark,
  title={Towards Efficient COVID-19 CT Annotation: A Benchmark for Lung and Infection Segmentation},
  author={Ma Jun and Wang Yixin and An Xingle and Ge Cheng and Yu Ziqi and Chen Jianan and Zhu Qiongjie and Dong Guoqiang and He Jian and He Zhiqiang and Ni Ziwei
```

```
and Yang Xiaoping},
  journal={arXiv preprint arXiv:2004.12537},
  year={2020}
}
```

- Jun Ma, Cheng Ge, Yixin Wang, Xingle An, Jiantao Gao, Ziqi Yu, Mingqing Zhang, Xin Liu, Xueyuan Deng, Shucheng Cao, Hao Wei, Sen Mei, Xiaoyu Yang, Ziwei Nie, Chen Li, Lu Tian, Yuntao Zhu, Qiongjie Zhu, Guoqiang Dong, Jian He. (2020). COVID-19 CT Lung and Infection Segmentation Dataset (Version 1.0) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3757476>

```
@dataset{COVID-19-CT-Seg-Dataset,
  author      = {Ma Jun and
                  Ge Cheng and
                  Wang Yixin and
                  An Xingle and
                  Gao Jiantao and
                  Yu Ziqi and
                  Zhang Mingqing and
                  Liu Xin and
                  Deng Xueyuan and
                  Cao Shucheng and
                  Wei Hao and
                  Mei Sen and
                  Yang Xiaoyu and
                  Nie Ziwei and
                  Li Chen and
                  Tian Lu and
                  Zhu Yuntao and
                  Zhu Qiongjie and
                  Dong Guoqiang and
                  He Jian},
  title       = {{COVID-19 CT Lung and Infection Segmentation
                  Dataset}},
  month       = Apr,
  year        = 2020,
  publisher   = {Zenodo},
  version     = {1.0},
  doi         = {10.5281/zenodo.3757476},
  url         = {https://doi.org/10.5281/zenodo.3757476}
}
```