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

Jun Ma, Yixin Wang, Xingle An, Cheng Ge, Ziqi Yu, Jianan Chen, Qiongjie Zhu, Guoqiang Dong, Jian He, Zhiqiang He, Ziwei Nie, Xiaoping Yang

Abstract—Accurate segmentation of lung and infection in COVID-19 CT scans plays an important role in the quantitative management of patients. Most of the existing studies are based on large and private annotated datasets that are impractical to obtain from a single institution, especially when radiologists are busy fighting the coronavirus disease. Furthermore, it is hard to compare current COVID-19 CT segmentation methods as they are developed on different datasets, trained in different settings, and evaluated with different metrics. In this paper, we created a COVID-19 3D CT dataset with 20 cases that contains 1800+ annotated slices and made it publicly available. To promote the development of annotation-efficient deep learning methods, we built three benchmarks for lung and infection segmentation that contain current main research interests, e.g., few-shot learning, domain generalization, and knowledge transfer. For a fair comparison among different segmentation methods, we also provide unified training, validation and testing dataset splits, and evaluation metrics and corresponding code. In addition, we provided more than 40 pre-trained baseline models for the benchmarks, which not only serve as out-of-the-box segmentation tools but also save computational time for researchers who are interested in COVID-19 lung and infection segmentation. To the best of our knowledge, this work presents the largest public annotated COVID-19 CT volume dataset, the first segmentation benchmark, and the most pre-trained models up to now. We hope these resources (https://gitee.com/junma11/COVID-19-CT-Seg-Benchmark) could advance the development of deep learning methods for COVID-19 CT segmentation with limited data.

Index Terms—COVID-19 CT, lung and infection segmentation, few-shot learning, domain generalization, knowledge transfer

I. INTRODUCTION

COVID-19 (novel coronavirus 2019) has spread all over the world during the past few months and caused over 170,000 deaths as of 23 April 2020[1]. It is projected to be a seasonal disease after the current outbreak [2], [3].

Jun Ma is with Department of Mathematics, Nanjing University of Science and Technology, Nanjing, 210094, P. R. China.

Yixin Wang is with Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Beijing, 100190, P. R. China

Xingle An is with CEC brain, Beijing, P. R. China

Cheng Ge is with Institute of Bioinformatics and Medical Engineering, Jiangsu University of Technology, Changzhou, 213001, P. R. China

Ziqi Yu is with Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai, P. R. China.

Jianan Chen is with Department of Medical Biophysics, University of Toronto, Toronto, ON, CA.

Qiongjie Zhu, Guoqiang Dong, and Jian He is with Department of Radiology, Nanjing Drum Tower Hospital, the Affiliated Hospital of Nanjing University Medical School, Nanjing, 210008, P. R. China.

Zhiqiang He is with Lenovo Ltd., Beijing, 100094, China.

Ziwei, Nie, and Xiaoping Yang (corresponding author) is with Department of Mathematics, Nanjing University, Nanjing, 210093, P. R. China.

Computed tomography (CT) is playing an important role in the fight against COVID-19 [4], [5], [6]. CT is shown to be more sensitive in the early diagnosis of COIVD-19 infection compared to Reverse Transcription-Polymerase Chain Reaction (RT-PCR) tests [7]. Wang et al. trained a deep learning model on 325 COVID-19 CT images and 740 typical pneumonia images. Their model identified 46 COVID-19 cases that were previously missed by the RT-PCR test. Further, quantitative information from CT images, such as lung burden, percentage of high opacity and lung severity score, can be used to monitor disease progression and help us understand the course of COVID-19 [8], [9].

Artificial Intelligence (AI) methods, especially deep learning-based methods, have been widely applied in medical image analysis to combat COVID-19 [10]. For example, AI can be used for building a contactless imaging workflow to prevent transmission from patients to health care providers [11]. In addition, most screening and segmentation algorithms for COVID-19 are developed with deep learning models[7], [12], [13], and most automatic diagnosis and infection quantification systems in COVID-19 studies rely on segmentation generated by deep neural networks[12], [14], [15], [16].

Although tremendous studies show that deep learning methods have potential for providing accurate and quantitative assessment of COVID-19 infection in CT images [10], the solutions mainly rely on large and private datasets. Due to the patient privacy and intellectual property issues, the datasets and solutions may not be publicly available. However, researchers may hope that the datasets, entire source code and trained models could be provided by authors [17].

Existing studies demonstrate that the classical U-Net (or V-Net) can achieve promising segmentation performance if hundreds of well-labelled training cases are available [15], [?]. 83.1% to 91.6% of segmentation performance in Dice score coefficient was reported in various U-Net-based approaches. Shan et al. developed a neural network based on V-Net and the bottle-neck structure to segment and quantify infection regions [?], [18]. In their human-in-the-loop strategy, they achieved 85.1%, 91.0% and 91.6% dice with 36, 114, and 249 labelled CT scans, respectively. Huang et al. [15] employed U-Net [19] for lung and infection using an annotated dataset of 774 cases, and demonstrated that trained model could be used for quantifying the disease burden and monitoring disease progression or treatment response. In general, we observed a trend that training deep learning models with more annotations will decrease the time needed for contouring the infection, and increase segmentation accuracy, which is consistent with

the Shan et al.'s findings [**?**]. However, annotations for 3D CT volume data are expensive to acquire because it not only relies on professional diagnosis knowledge of radiologists, but also takes much time and labor, especially in current situation. Thus, a critical question would be:

how can we annotate COVID-19 CT scans efficiently?

Towards this question, three basic but important problems remain unsolved:

- There is no publicly well-labelled COVID-19 CT 3D dataset; Data collection is the first and essential step to develop deep learning methods for COVID-19 segmentation. Most of the existing studies rely on private and large dataset with hundreds of annotated CT scans.
- There are no public benchmarks to evaluate different deep learning-based solutions and different studies use various evaluation metrics. Similarly, datasets for training, validating and testing are split diversely, which also makes it hard for readers to compare those methods. For example, although MedSeg (http://medicalsegmentation.com/covid19/) dataset with 100 annotated slices were used in [20], [21], [22], [23] to develop COVID-19 segmentation methods, it was split in different ways and the developed methods were evaluated by different metrics.
- There are no publicly available baseline models trained for COVID-19. U-Net, a well-known segmentation architecture, is commonly used as a baseline network in many studies. However, due to different implementations, the reported performance variesin different studies [21], [24], even though the experiments are conducted on the same dataset.

In this paper, we focus on annotation-efficient deep learning solutions and aim to alleviate the above problems by providing a well-labelled COVID-19 CT dataset and a benchmark. In particular, we first annotated left lung, right lung and infections of 20 COVID-19 3D CT scans, and then established three tasks to benchmark different deep learning strategies with limited training cases. Finally, we built baselines for each task based on U-Net.

Our tasks target on three popular research fields in medical imaging community:

- Few-shot learning: building high performance models from very few samples. Although most existing approaches focus on natural images, it has received growing attention in medical imaging because generating labels for medical images is much more difficult [25], [26].
- Domain generalization: learning from known domains and applying to an unknown target domain that has different data distribution [27]. Adversarial training is usually adopted to achieve it in recent studies. [28], [29].
- Knowledge transfer: reusing existing annotations to boost the training/fine-tuning on a new dataset. Transferring knowledge from large public datasets is a well-established practice in medical imaging analysis [30], [31]. Using generative adversarial network to transfer knowledge from publicly available annotated datasets to new datasets

for segmentation also receives significant attention [32], [33].

Our contributions can be summarized as following:

- 20 well-labelled COVID-19 CT volume data are publicly released to the community. To the best of our knowledge, this is the largest public COVID-19 3D CT segmentation datasets.
- 3 benchmark tasks are set up to promote studies on annotation-efficient deep learning segmentation for COVID-19 CT scans. Specifically, we focus on few-shot learning, domain generalization, and knowledge transfer.
- 40+ trained models are publicly available that can serve as baselines and out-of-the-box tools for COVID-19 CT lung and infection segmentation.

II. DATASETS

Annotations of COVID-19 CT scans are scarce, but several lung CT annotations with other diseases are publicly available. Thus, one of the main goals for our benchmark is to explore these existing annotations to assist COVID-19 CT segmentation. This section introduces the public datasets used in our segmentation benchmarks. Figure 1 presents some examples from each dataset.

A. Existing lung CT segmentation datasets

1) StructSeg lung organ segmentation: 50 lung cancer patient CT scans are accessible, and all the cases are from one medical center. This dataset served as a segmentation challenge[1] during MICCAI 2019. Six organs are annotated, including left lung, right lung, spinal cord, esophagus, heart, and trachea. In this paper, we only use the left lung and right lung annotations.

2) NSCLC left and right lung segmentation: This dataset consists of left and right thoracic volume segmentations delineated on 402 CT scans from The Cancer Imaging Archive NSCLC Radiomics [34], [35], [36].

B. Existing lung lesion CT segmentation datasets

1) MSD Lung tumor segmentation: This dataset is comprised of patients with non-small cell lung cancer from Stanford University (Palo Alto, CA, USA) publicly available through TCIA. The dataset served as a segmentation challenge[2] during MICCAI 2018. The tumor is annotated by an expert thoracic radiologist, and 63 labelled CT scans are available.

2) StructSeg Gross Target Volume segmentation of lung cancer: The same 50 lung cancer patient CT scans as the above StructSeg lung organ segmentation dataset are provided, and gross target volume of tumors are annotated in each case.

3) NSCLC Pleural Effusion (PE) segmentation: The CT scans in this dataset are the same as those in NSCLC left and right lung segmentation dataset, while pleural effusion is delineated for 78 cases [34], [35], [36].

---

[1]MICCAI 2019 StructSeg: https://structseg2019.grand-challenge.org
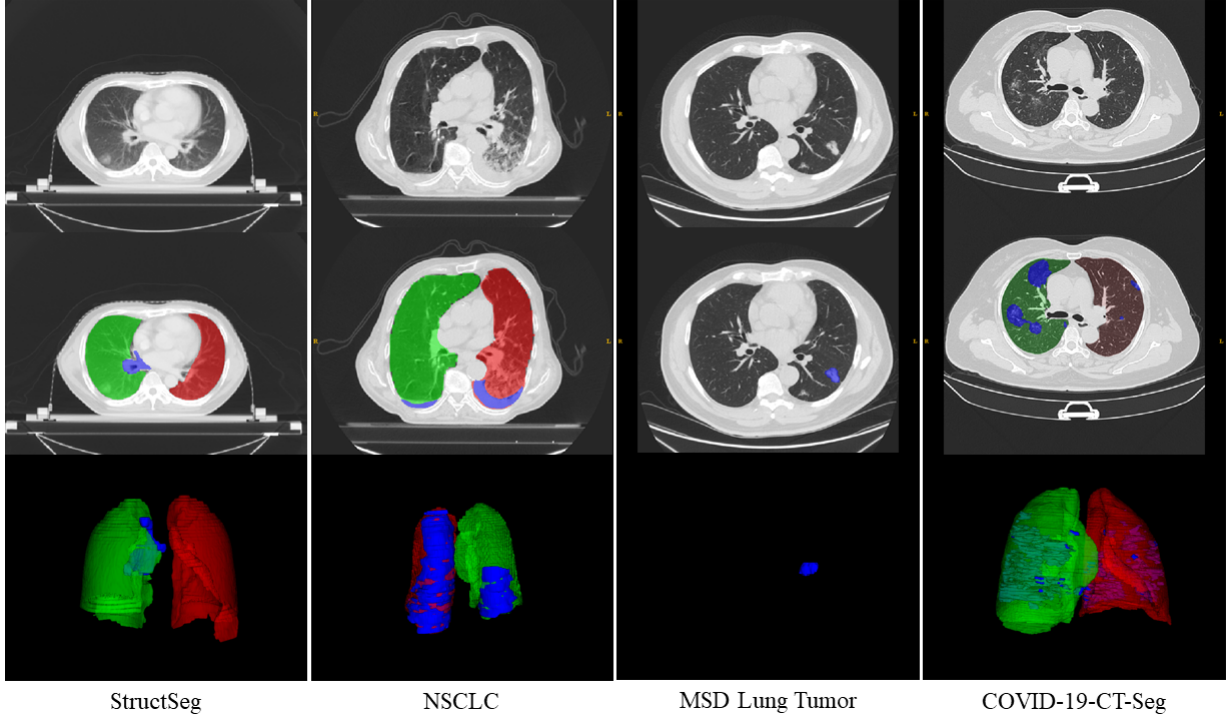[2]MICCAI 2018 MSD: http://medicaldecathlon.com/

Fig. 1. Examples of four lung CT datasets. The 1st and 2nd row denote original CT images and corresponding ground truth of lung and lesions, respectively. The 3rd row shows the 3D rendering results of ground truth. MSD Lung Tumor dataset (3rd column) does not provide lung masks.

C. Our COVID-19-CT-Seg dataset

We collected 20 public COVID-19 CT scans from the Coronacases Initiative and Radiopaedia, which can be freely downloaded[3] with CC BY-NC-SA license. The left lung, right lung, and infection were firstly delineated by junior annotators, then refined by two radiologists with 5 years experience, and finally all the annotations were verified and refined by a senior radiologist with more than 10 years experience in chest radiology. On average, it takes about 400 ± 45 minutes to delineate one CT scan with 250 slices. We have made all the annotations publicly available [37] at https://zenodo.org/record/3757476 with CC BY-NC-SA license.

III. BENCHMARK SETTINGS FOR COVID-19 CT LUNG AND INFECTION SEGMENTATION

As mentioned in Section 1, there is a need for innovative strategies that enable annotation-efficient methods for COVID-CT segmentation. Thus, we set up three tasks to evaluate potential annotation-efficient strategies. In particular, we focus on learning to segment left lung, right lung and infection in COVID-19 CT scans using

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

[3]https://github.com/ieee8023/covid-chestxray-dataset

A. Task 1: Learning with limited annotations

Task 1 is designed to address the problem of few-short learning, where few annotations are available for training. This task is based on the COVID-19-CT-Seg dataset only. It contains three subtasks aiming to segment lung, infection and both of them, respectively. For each subtask, 5-fold cross validation results (based on a pre-defined dataset split file) are reported. In each fold, 4 training cases are used for training and the rest 16 cases are used for testing (Table I).

TABLE I
EXPERIMENTAL SETTINGS OF TASK 1 (LEARNING WITH LIMITED ANNOTATIONS) FOR LUNG AND INFECTION SEGMENTATION IN COVID-19 CT SCANS. ALL THE EXPERIMENTS ARE BASE ON THE COVID-19-CT-SEG DATASET.

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

B. Task 2: Learning to segment COVID-19 CT scans from non-COVID-19 CT scans

Task 2 is designed to address the problem of domain generalization, where only out-of-domain data are available for training. Specifically, in the first subtask, the StructSeg Lung dataset and the NSCLC Lung dataset are used for training a lung segmentation model. In the second subtask, the MSD Lung Tumor, the StructSeg Gross Target and the NSCLC Pleural Effusion datasets are used. For both subtasks, 80% of the data are randomly selected for training and the rest 20%

are held-out for validation. All labelled COVID-CT-Seg CT scans are kept for testing (Table II).

C. Task 3: Learning with both COVID-19 and non-COVID-19 CT scans

Task 3 is designed to address the problem of knowledge transfer with heterogeneous annotations, where out-of-domain data are used to boost the annotation-efficient training on limited in-domain data. Specifically, in both subtasks (lung segmentation and lung infection segmentation), 80% and 20% of out-of-domain data are used for training and validation, while 20% and 80% of in-domain data are used for training and testing, respectively (Table III).

D. Evaluation metrics

Motivated by the evaluation methods of the well-known medical image segmentation decathlon[4], we also employ two complementary metrics to evaluate the segmentation performance. Dice similarity coefficient, a region-based measure, is used to evaluate the region overlap. Normalized surface Dice [38], a boundary-based measure is used to evaluate how close the segmentation and ground truth surfaces are to each other at a specified tolerance τ. For both two metrics, higher scores admit better segmentation performance, and 100% means perfect segmentation. Let G, S denote ground truth and segmentation, respectively. We formulate the definitions of the two measures as follows:

1) Region-based measure:

$$DSC(G, S) = \frac{2|G \cap S|}{|G| + |S|};$$

2) Boundary-based measure:

$$NSD(G, S) = \frac{|\partial G \cap B_{\partial S}^{(\tau)}| + |\partial S \cap B_{\partial G}^{(\tau)}|}{|\partial G| + |\partial S|}$$

where $B_{\partial G}^{(\tau)}, B_{\partial S}^{(\tau)} \subset R^3$ denote the border region of ground truth and segmentation surface at tolerance τ, which are defined as $B_{\partial G}^{(\tau)} = \{x \in R^3 \,|\, \exists \tilde{x} \in \partial G, \|x - \tilde{x}\| \leq \tau\}$ and $B_{\partial S}^{(\tau)} = \{x \in R^3 \,|\, \exists \tilde{x} \in \partial S, \|x - \tilde{x}\| \leq \tau\}$, respectively. We set tolerance τ as $1mm$ and $3mm$ for lung segmentation and infection segmentation, respectively. The tolerance is computed by measuring the inter-rater segmentation variation between two different radiologists, which is also in accordance with another independent study [38].

The main benefit of introducing Surface Dice is that it ignores small boundary deviations because small inter-observer errors are also unavoidable and often not clinically relevant when segmenting the objects by radiologists.

E. U-Net baselines: oldies but goldies

U-Net ([19], [39]) has been proposed for 5 years, and many variants have been proposed to improve it. However, recent study [40] demonstrates that a basic U-Net is still difficult to surpass if the corresponding pipeline is designed adequately.

[4]http://medicaldecathlon.com/files/MSD-Ranking-scheme.pdf

In particular, nnU-Net (no-new-U-Net) [40] was proposed to automatically adapt preprocessing strategies and network architectures (i.e., the number of pooling, convolutional kernel size, and stride size) to a given 3D medical dataset. Without manual tuning, nnU-Net achieved better performance than most specialised deep learning pipelines in 19 public international segmentation competitions and set a new state-of-the-art in the majority of 49 tasks. The source code is publicly available at https://github.com/MIC-DKFZ/nnUNet.

U-Net is often used as a baseline model in existing COVID-19 CT segmentation studies. However, reported results vary a lot even in the same dataset, which make it hard to compare different studies. To standardize the U-Net performance, we build our baselines on nnU-Net, which is the most powerful U-Net implementation to the best of our knowledge. To make it comparable between different tasks, we manually adjust the patch size and network architectures in task 2 and task 3 to be the same as task 1. Figure 2 shows details of the U-Net architecture.

IV. RESULTS AND DISCUSSION

A. Results of Task 1: Learning with limited annotations

Table IV presents average DSC and NSD results of lung and infection of each subtask in Task 1. It can be found that

- the average DSC and NSD values among different folds vary greatly. It demonstrates that reporting 5-fold cross validation results is necessary to obtain an object evaluation because 1-fold results may be biased.
- promising results for left and right lung segmentation in COVID-19 CT scans can be achieved with as few as four training cases. Models trained for segmenting lung obtained significantly better results compared with those trained for segmenting lung and infection simultaneously.
- There is still large room for improving infection segmentation with limited annotations. Statistical tests are needed to justify whether there is an improvement from training the model for segmenting lung and infection at the same time.

B. Results of Task 2: Learning to segment COVID-19 CT scans from non-COVID-19 CT scans

This task is quite challenging as the model does not see any cases from target domain during training. In other words, the trained models are expected to generalize to the unseen domain (COVID-19 CT). Table V shows left lung and right lung segmentation results in terms of average and standard deviation values of DSC and NSD. It can be found that

- 3D U-Net achieves excellent performance in terms of DSC on the validation set that is from the same source domain. Average NSD values are lower than DSC values, implying that most of the errors come from the boundary.
- the performance on the testing set drops significantly on both subtasks. The performance of the model trained on NSCLC Lung dataset is worse than the model trained on StuctSeg lung. We suppose it is because the difference in the distribution lung appearance is smaller between StructSeg and COVID-19-CT.

TABLE II
Experimental settings of Task 2 (Learning to segment COVID-19 CT scans from non-COVID-19 CT scans) for lung and infection segmentation in COVID-19 CT scans.

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

TABLE III
Experimental settings of Task 3 (Learning with both COVID-19 and non-COVID-19 CT scans) for lung and infection segmentation in COVID-19 CT scans.

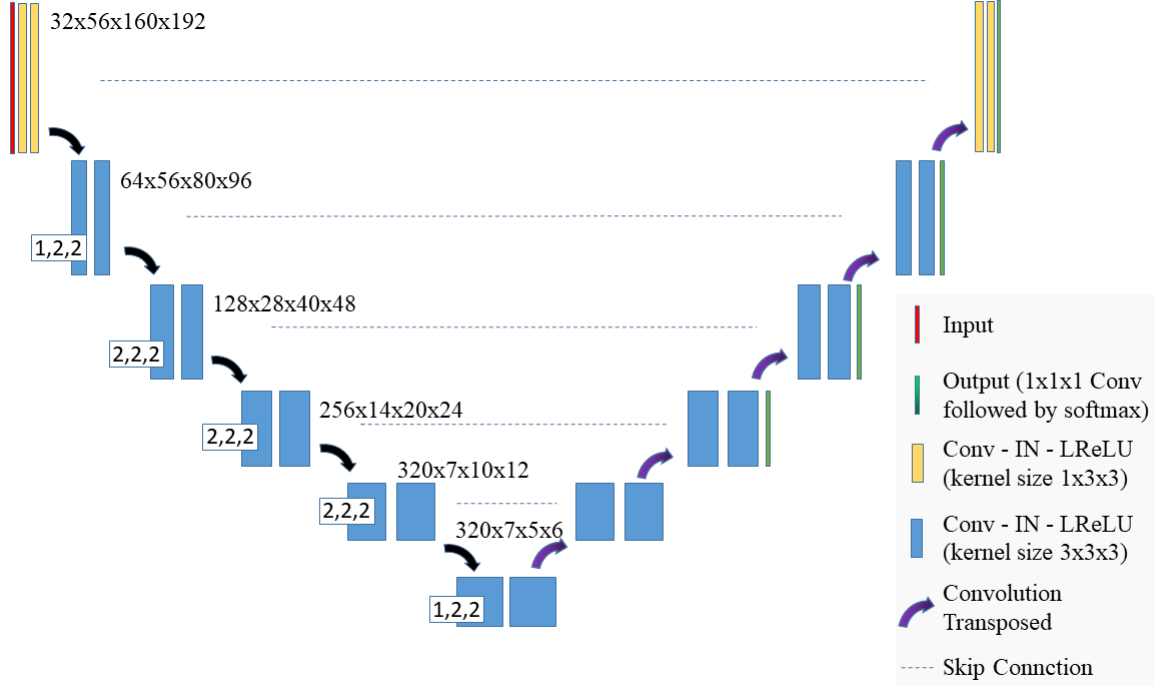| Seg. Task	Training		Validation	Testing
Lung	StructSeg Lung (40) NSCLC Lung (322)	COVID-CT-Seg Lung (4)	StructSeg Lung (10) NSCLC Lung (80)	COVID-CT-Seg Lung (16)
Infection	MSD Lung Tumor (51) StructSeg Gross Target (40) NSCLC Pleural Effusion (62)	COVID-CT-Seg Infection (4)	MSD Lung Tumor (12) StructSeg Gross Target (10) NSCLC Pleural Effusion (16)	COVID-CT-Seg Infection (16)

Fig. 2. Details of the 3D U-Net architecture that is used in this work.

Table VI shows quantitative infection segmentation results in terms of average and standard deviation values of DSC and NSD. It can be found that

- on the validation set, the performance of lesion segmentation is not as good as the performance of lung segmentation (Table V), which means that tumor segmentation remains a challenging problem. This observation is in line with recent results in MICCAI tumor segmentation challenge, i.e. liver tumor segmentation [41] and kidney tumor segmentation [42].
- The models almost fail to predict COVID-19 infections on testing set, which highlights that the lesion appearances differ significantly among lung cancer, pleural effusion, and COVID-19 infections in CT scans.

C. Limitation

One possible limitation would be the number of cases in our dataset is relatively small. On one hand, we focus on how to learn from a few training cases(task 1) and how to efficiently use existing annotations. Thus, the number of training cases is acceptable for our benchmark tasks. On the other hand, we have *16 or 20* testing cases in each task. The number of testing cases is comparable with the setting of recent MICCAI 2020 segmentation challenges. For example, MyoPS 2020 (Myocardial pathology segmentation combining multi-sequence CMR) has *15* testing cases [43]. StructSeg (Automatic Structure Segmentation for Radiotherapy Planning Challenge 2020) has *10* testing cases [44] and ASOCA (Automated Segmentation

TABLE IV

QUANTITATIVE RESULTS OF 5-FOLD CROSS VALIDATION OF TASK 1: LEARNING WITH LIMITED ANNOTATIONS. FOR EACH FOLD, AVERAGE DSC AND NSD VALUES ARE REPORTED. THE LAST ROW SHOWS THE AVERAGE RESULTS OF 80 (= 5 FOLDS ×16 TESTING CASES PER FOLD) TESTING CASES.

Subtask	Lung				Infection		Lung and Infection					
	Left Lung		Right Lung				Left Lung		Right Lung		Infection	
	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD
Fold-0	84.98.2	68.713.3	85.213.0	70.615.8	68.120.5	70.921.3	53.828.4	39.118.3	65.519.4	47.414.3	65.423.9	068.223.2
Fold-1	80.314.5	61.815.1	83.99.6	68.39.0	71.320.5	71.823.0	40.318.7	27.512.0	60.111.1	41.79.9	64.721.8	60.625.1
Fold-2	87.112.1	74.316.0	90.38.2	78.512.0	66.221.7	71.724.2	80.318.8	66.818.8	85.212.4	68.615.1	60.727.6	62.528.9
Fold-3	88.47.0	75.28.8	89.96.3	78.58.0	68.123.1	70.827.1	79.713.6	65.414.4	84.09.8	67.713.0	62.027.9	65.328.9
Fold-4	88.376.0	75.811.0	90.27.0	78.310.2	62.726.9	64.928.2	72.421.1	58.620.8	80.913.4	63.415.9	51.430.2	51.931.0
Avg	**85.810.5**	**71.213.8**	**87.99.3**	**74.811.9**	**67.322.3**	**70.024.4**	65.425.6	51.622.9	75.316.8	57.917.5	60.826.3	61.627.5

TABLE V

QUANTITATIVE RESULTS (MEAN ± STANDARD DEVIATION) OF LUNG SEGMENTATION IN TASK 2.

Subtask	Validation Set				(Unseen) Testing Set			
	Left Lung		Right Lung		Left Lung		Right Lung	
	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD
StructSeg Lung	96.4 ± 1.4	74.6 ± 9.1	97.3 ± 0.3	74.3 ± 7.2	92.2 ± 19.7	82.0 ± 15.7	95.5 ± 7.2	84.2 ± 11.6
NSCLC Lung	95.3 ± 4.9	80.2 ± 8.3	95.4 ± 10.9	80.7 ± 10.7	57.5 ± 21.5	46.9 ± 16.9	72.2 ± 15.3	51.7 ± 16.8

TABLE VI

QUANTITATIVE RESULTS (MEAN ± STANDARD DEVIATION) OF INFECTION SEGMENTATION IN TASK2.

Subtask	Validation Set		(Unseen) Testing Set	
	DSC	NSD	DSC	NSD
MSD Lung Tumor	67.2 ± 27.1	77.1 ± 31.4	25.2 ± 27.4	26.0 ± 28.5
StructSeg Tumor	71.3 ± 29.6	70.3 ± 29.5	6.0 ± 12.7	5.5 ± 10.7
NSCLC-PE	64.4 ± 45.5	73.7 ± 12.9	0.4 ± 0.8	3.7 ± 4.8

Of Coronary Arteries) has *20* testing cases [45].

Another possible limitation is that only quantitative results of 3D U-Net are reported. We will also provide 2D U-Net baselines for all tasks in the near future.

V. CONCLUSION

In this paper, we created a COVID-19 CT dataset, established three segmentation benchmark tasks towards annotation-efficient methods, and provided 40+ baselines models based on state-of-the-art segmentation architectures. All the related results are publicly available at https://gitee.com/junma11/COVID-19-CT-Seg-Benchmark. We hope this work can help community to accelerate designing and developing AI-based COVID-19 CT analysis tools with limited data.

REFERENCES

[1] W. H. Organization *et al.*, "Coronavirus disease 2019 (covid-19): situation report, 94," 2020.

[2] R. A. Neher, R. Dyrdak, V. Druelle, E. B. Hodcroft, and J. Albert, "Potential impact of seasonal forcing on a sars-cov-2 pandemic," *Swiss Medical Weekly*, vol. 150, no. 1112, 2020.

[3] S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, and M. Lipsitch, "Projecting the transmission dynamics of sars-cov-2 through the postpandemic period," *Science*, 2020.

[4] A. M. Foust, G. S. Phillips, W. C. Chu, P. Daltro, K. M. Das, P. Garcia-Pea, T. Kilborn, A. J. Winant, and E. Y. Lee, "International expert consensus statement on chest imaging in pediatric covid-19 patient management: Imaging findings, imaging study reporting and imaging study recommendations," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200214, 2020.

[5] M. Oudkerk, H. R. Bller, D. Kuijpers, N. van Es, S. F. Oudkerk, T. C. McLoud, D. Gommers, J. van Dissel, H. ten Cate, and E. J. van Beek, "Diagnosis, prevention, and treatment of thromboembolic complications in covid-19: Report of the national institute for public health of the netherlands," *Radiology*, vol. 0, no. 0, p. 201629, 2020.

[6] P. J. Mazzone, M. K. Gould, D. A. Arenberg, A. C. Chen, H. K. Choi, F. C. Detterbeck, F. Farjah, K. M. Fong, J. M. Iaccarino, S. M. Janes, J. P. Kanne, E. A. Kazerooni, H. MacMahon, D. P. Naidich, C. A. Powell, S. Raoof, M. P. Rivera, N. T. Tanner, L. K. Tanoue, A. Tremblay, A. Vachani, C. S. White, R. S. Wiener, and G. A. Silvestri, "Management of lung nodules and lung cancer screening during the covid-19 pandemic: Chest expert panel report," *Radiology: Imaging Cancer*, vol. 2, no. 3, p. e204013, 2020.

[7] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest ct for covid-19: comparison to rt-pcr," *Radiology*, p. 200432, 2020.

[8] G. Chassagnon, M. Vakalopoulou, E. Battistella, S. Christodoulidis, T.-N. Hoang-Thi, S. Dangeard, E. Deutsch, F. Andre, E. Guillo, N. Halm, S. El Hajj, F. Bompard, S. Neveu, C. Hani, I. Saab, A. Campredon, H. Koulakian, S. Bennani, G. Freche, A. Lombard, L. Fournier, H. Monnier, T. Grand, J. Gregory, A. Khalil, E. Mahdjoub, P.-Y. Brillet,

---

[5] https://github.com/HzFu/COVID19_imaging_AI_paper_list
[6] http://medicalsegmentation.com/covid19/

S. Tran Ba, V. Bousson, M.-P. Revel, and N. Paragios, "Ai-driven ct-based quantification, staging and short-term outcome prediction of covid-19 pneumonia," *medRxiv*, 2020.

[9] S. Chaganti, A. Balachandran, G. Chabin, S. Cohen, T. Flohr, B. Georgescu, P. Grenier, S. Grbic, S. Liu, F. Mellot, *et al.*, "Quantification of tomographic patterns associated with covid-19 from chest ct," *arXiv preprint arXiv:2004.01279*, 2020.

[10] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," *IEEE Reviews in Biomedical Engineering*, 2020.

[11] Y. Wang, X. Lu, Y. Zhang, X. Zhang, K. Wang, J. Liu, X. Li, R. Hu, X. Meng, S. Dou, *et al.*, "Precise pulmonary scanning and reducing medical radiation exposure by developing a clinically applicable intelligent ct system: Toward improving patient care," *EBioMedicine*, vol. 54, p. 102724, 2020.

[12] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for covid-19 from chest ct using weak label," *medRxiv*, 2020.

[13] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, *et al.*, "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, p. 200905, 2020.

[14] Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, and H. Shi, "Longitudinal assessment of covid-19 using a deep learning–based quantitative ct pipeline: Illustration of two cases," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200082, 2020.

[15] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, and L. Xia, "Serial quantitative chest ct assessment of covid-19: Deep-learning approach," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200075, 2020.

[16] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, Y. Cha, W. Liang, C. Wang, et al., and G. Wang, "Clinically applicable ai system for accurate diagnosis, quantitative measurements and prognosis of covid-19 pneumonia using computed tomography," *Cell*, 2020.

[17] A. M. V. Dadário, J. P. Q. Paiva, R. C. Chate, B. S. Machado, and G. Szarf, "Regarding "artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct"," *Radiology*, p. 201178, 2020.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.

[20] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct scans," *arXiv*, 2020.

[21] T. Zhou, S. Canu, and S. Ruan, "An automatic covid-19 ct segmentation network using spatial and channel attention mechanism," *arXiv preprint arXiv:2004.06673*, 2020.

[22] Q. Yu, Y. Liu, and J. Xu, "Miniseg: An extremely minimum network for efficient covid-19 segmentation," *arXiv preprint arXiv:2004.09750*, 2020.

[23] A. Amyar, R. Modzelewski, and S. Ruan, "Multi-task deep learning based ct imaging analysis for covid-19: Classification and segmentation," *medRxiv*, 2020.

[24] X. Chen, L. Yao, and Y. Zhang, "Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images," *arXiv preprint arXiv:2004.05645*, 2020.

[25] A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca, "Data augmentation using learned transformations for one-shot medical image segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[26] A. G. Roy, S. Siddiqui, S. Pölsterl, N. Navab, and C. Wachinger, "squeeze & exciteguided few-shot segmentation of volumetric images," *Medical image analysis*, vol. 59, p. 101587, 2020.

[27] D. Li, J. Zhang, Y. Yang, C. Liu, Y.-Z. Song, and T. M. Hospedales, "Episodic training for domain generalization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 1446–1455.

[28] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[29] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, "Diverse image-to-image translation via disentangled representations," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[30] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging*, vol. 3, no. 3, p. 034501, 2016.

[31] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[32] F. Yu, J. Zhao, Y. Gong, Z. Wang, Y. Li, F. Yang, B. Dong, Q. Li, and L. Zhang, "Annotation-free cardiac vessel segmentation via knowledge transfer from retinal images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 714–722.

[33] D. Eschweiler, T. Klose, F. N. Müller-Fouarge, M. Kopaczka, and J. Stegmaier, "Towards annotation-free segmentation of fluorescently labeled cell membranes in confocal microscopy images," in *International Workshop on Simulation and Synthesis in Medical Imaging*, 2019, pp. 81–89.

[34] K. Kiser, S. Ahmed, S. Stieb, A. Mohamed, H. Elhalawani, P. Park, N. Doyle, B. Wang, A. Barman, C. Fuller, and L. Giancardo, "Data from thethoracic volume and pleural effusion segmentations in diseased lungs for benchmarking chest ct processing pipelines," *The Cancer Imaging Archive*, 2020.

[35] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, no. 1, pp. 1–9, 2014.

[36] K. W. Clark, B. A. Vendt, K. E. Smith, J. B. Freymann, J. S. Kirby, P. Koppel, S. M. Moore, S. R. Phillips, D. R. Maffitt, M. Pringle, L. Tarbox, and F. W. Prior, "The cancer imaging archive (tcia): Maintaining and operating a public information repository," *J. Digital Imaging*, no. 6, pp. 1045–1057, 2013.

[37] J. Ma, C. Ge, Y. Wang, X. An, J. Gao, Z. Yu, M. Zhang, X. Liu, X. Deng, S. Cao, H. Wei, S. Mei, X. Yang, Z. Nie, C. Li, L. Tian, Y. Zhu, Q. Zhu, G. Dong, and J. He, "COVID-19 CT Lung and Infection Segmentation Dataset," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3757476

[38] S. Nikolov, S. Blackwell, R. Mendes, J. De Fauw, C. Meyer, C. Hughes, H. Askham, B. Romera-Paredes, A. Karthikesalingam, C. Chu, *et al.*, "Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy," *arXiv preprint arXiv:1809.04430*, 2018.

[39] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, 2016, pp. 424–432.

[40] F. Isensee, J. Petersen, S. A. A. Kohl, P. F. Jäger, and K. H. Maier-Hein, "nnu-net: Breaking the spell on successful medical image segmentation," *arXiv preprint arXiv:1904.08128*, 2020.

[41] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser, *et al.*, "The liver tumor segmentation benchmark (lits)," *arXiv preprint arXiv:1901.04056*, 2019.

[42] 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," *arXiv preprint arXiv:1912.01054*, 2019.

[43] X. Zhuang and L. Li, "Multi-sequence CMR based myocardial pathology segmentation challenge," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3715932

[44] H. Li and M. Chen, "Automatic Structure Segmentation for Radiotherapy Planning Challenge 2020," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3718885

[45] R. Gharleghi, D. G. Samarasinghe, P. A. Sowmya, and D. S. Beier, "Automated segmentation of coronary arteries," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3714986