2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Assessing Lobe-wise Burden of COVID-19 Infection in Computed Tomography of Lungs using Knowledge Fusion from Multiple Datasets

Mahalakshumi Visvanathan[1], Velmurugan Balasubramanian[2], Rachana Sathish[2], Suhasini Balasubramaniam[3] and Debdoot Sheet[2]

***Abstract*— Segmentation of COVID-19 infection in the lung tissue and its quantification in individual lobes is pivotal to understanding the disease's effect. It helps to determine the disease progression and gauge the extent of medical support required. Automation of this process is challenging due to the lack of a standardized dataset with voxel-wise annotations of the lung field, lobes, and infections like ground-glass opacity (GGO) and consolidation. However, multiple datasets have been found to contain one or more classes of the required annotations. Typical deep learning-based solutions overcome such challenges by training neural networks under adversarial and multi-task constraints. We propose to train a convolutional neural network to solve the challenge while it learns from multiple data sources, each of which is annotated for only a few classes. We have experimentally verified our approach by training the model on three publicly available datasets and evaluating its ability to segment the lung field, lobes and COVID-19 infected regions. Additionally, eight scans that previously had annotations for infection and lung have been annotated for lobes. Our model quantifies infection per lobe in these scans with an average error of 4.5%.**

I. INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a communicable disease caused by SARS-COV-2, which spreads through infected droplets, causing lesions predominantly in the lung parenchyma of affected individuals. A study with more than $1,000$ patient cohorts reported that the CT based assessment of the distribution of these lesions across different lung lobes has a diagnostic sensitivity of 97%, the positive predictive value of 65%, and a negative predictive value of 83%[1]. CT based clinical severity assessment has also been suggested by [2], which involves visual quantification of the extent of the involvement of these lesions across different lobes. We aim to detect and objectively quantify two of the most commonly occurring lesions, namely GGO and consolidation across different lobes, to be helpful for diagnosis and severity assessment.

A. Challenge

At the time of submitting this publication, there were no publicly available datasets of lung CT scans with annotations for both COVID infection and lung lobe segments needed to train a deep neural network (DNN) for lobe wise quantification. Thus, the challenge lies in learning from multiple data sources with different sets of annotations.

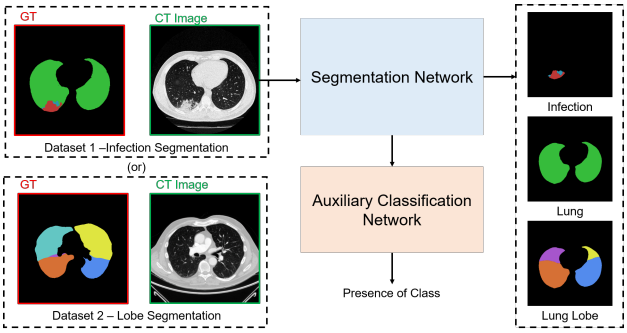


Fig. 1. Overview of our method of knowledge fusion.

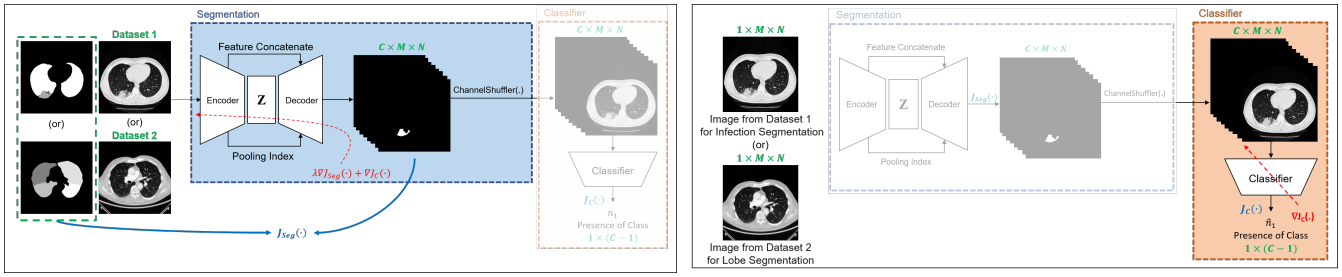
B. Approach

We propose a 2D convolutional neural network (CNN) based learning framework that can segment GGO, consolidation, lungs and the lobes from CT scan slices. It consists of a segmentation network and auxiliary classification network that aids the learning process, as shown in Fig. 1. We use two publicly available datasets containing annotations for infection and lung and another dataset with annotations for the lobes.

C. Related Works

3D-CNN [3] and relational two-stage U-net [4] have been used for lung lobe segmentation in CT images. The task of COVID-19 infection segmentation has also been performed using 2D and 3D UNet architecture by [5]. Segmentation of lung lobe and COVID-19 infection segmentation together was attempted by [6] using two separate networks to do the segmentation; a relational two-stage U-Net architecture as proposed by [4] was used for lobe segmentation, and a 3D U-net was used to segment GGO and Consolidation. [7] proposed a convolutional neural network with two branches, one for classification of disease (COVID-19/Normal) and another for lesion segmentation. A key feature of this method is the lesion attention (LA) module which combines features of the classification and segmentation branches of the network. To facilitate learning from noisy labels [8] designed a 2D CNN model called COPLE-Net, which is trained using a noise-robust Dice loss. The proposed framework also leverages self-ensembling to deal with noisy labels, where a teacher network is updated as an exponential moving average of a

[1]Mahalakshumi Visvanathan is with Sri Venkateswara College of Engineering, Chennai, India
[2]Rachana Sathish and Debdoot Sheet are with the Indian Institute of Technology Kharagpur, India. Velmurugan Balasubramanian was with Indian Institute of Technology Kharagpur, India when this work was done. debdoot@ee.iitkgp.ac.in
[3]Suhasini Balasubramaniam is with Government Medical College, Omandurar Government Estate, Chennai, India



(a) Training of the segmentation network

(b) Training of the auxillary classifier

Fig. 2. Training of our proposed framework consisting of the two different networks.

student network. COVID-19 lesion segmentation along with disease severity prediction is addressed in the work by [9]. We adopt an approach similar to [10] for knowledge fusion from multiple datasets.

II. Problem Statement

The task of infection, lung and lobe segmentation in CT scans assigns each pixel in a 2D axial slice of CT with an appropriate set of labels from the following nine classes: background, GGO, Consolidation, Lung, left upper (LU) lobe, left lower (LL) lobe, right upper (RU) lobe, right middle (RM) lobe and right lower (RL) lobe. Since annotations for all these classes are not available in a single dataset, the task at hand involves learning from multiple datasets with partial annotations. Also, the learning process must ensure that knowledge aggregated from one dataset is retained when trained on other datasets.

III. Method

A. Segmentation Network

The encoder-decoder segmentation network in Fig. 2(a) follows the architecture proposed by [11]. It has feature concatenation across the matched layers and index passing based un-pooling at the decoder. We added batch normalization after every convolutional layer. The final output of the segmentation network consists of C channels, where the first channel represents the background and the remaining $C-1$ channels represent the classes to be segmented.

B. Auxiliary Classification Network

This network consists of 7 convolutional layers, each with a kernel size of 3×3 (Fig. 2(b)). First 6 convolutional layers are followed by $LeakyRelu()$ non-linearity, batch normalization and max-pooling operation with a kernel size of 3×3. The final convolutional layer contains $C-1$ channels and is followed by a sigmoid activation function. The output of this network, $\hat{n_1}$, is a multi-hot encoded vector indicating the presence of a class in the input image.

C. Training methodology

In one training iteration, the model is trained on one mini-batch of each of the infection segmentation data and lobe segmentation data.

Phase 1: The input image $\mathcal{I}$ is passed through the segmentation CNN $Net_{Seg}(\cdot)$ and the output $\hat{O}$ of size $C \times M \times N$ is obtained. Masked binary cross-entropy loss $J_{Seg}(\cdot)$ is computed as,

$$J_{Seg}(\hat{O}, O) = BCE(m_{\mathcal{I}} \cdot \hat{O}, O) \tag{1}$$

where O is the ground-truth (GT) annotation, $m_{\mathcal{I}}$ is a multi-hot vector indicating annotated classes for $\mathcal{I}$.

Phase 2: $C-1$ channels of $\hat{O}$ excluding the background are shuffled and appended with the input image and fed into the auxiliary classification network $net_{class}(\cdot)$. A cross-entropy loss $J_C(\cdot)$ is calculated between the predicted classes $\hat{n_1}$ and the ground truth n_1. The gradient ∇J_C is used to update the classifier, as shown in Fig. 2(b).

Phase 3: The combined loss J_{total} is then used to update the segmentation network, as shown in Fig. 2(a).

$$J_{total} = \lambda J_{Seg}(\cdot) + J_C(\cdot) \tag{2}$$

IV. Experiments

A. Dataset description

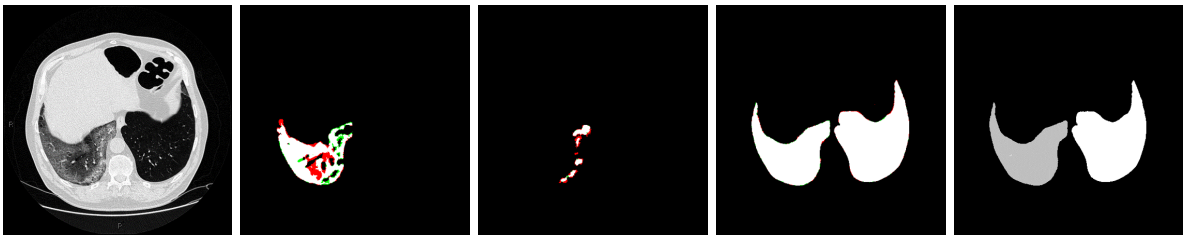
Lobe segmentation: We use the dataset (*Dset1*) provided by [3], which has annotations for five lung lobes (right upper, right middle lobe, right lower lobe, left upper lobe and left lower lobe). *Infection segmentation:* We use two datasets - CNCB dataset[1] (*Dset2*) and MedSeg dataset[2] (*Dset3*), with annotations for lungs, Ground glass opacities (GGO) and consolidation. MedSeg dataset has a total of 829 annotated slices corresponding to 9 volumes and the CNCB dataset has annotations for a total of 750 slices corresponding to 150 patients.

Generalization of segmentation: In order to test the generalization capability of the proposed model, we use another dataset from Mosmed.ai[3] (*Dset4*) having annotations for GGO and consolidation in 50 patients with around 40 slices present per patient. The proposed model and baselines are not trained on this dataset.

B. Baselines

- **BL1:** 3D CNN proposed by [3] for lobe segmentation. To ensure fair comparison we have used the same data split of Dset1 in our experiments.

[1] http://ncov-ai.big.ac.cn/download?lang=en
[2] http://medicalsegmentation.com/covid19/
[3] https://mosmed.ai/en/datasets/



(a) CT image slice (b) GGO (0.91) (c) Consolidation (0.78) (d) Lung field (0.99) (e) Lobes

Fig. 3. Figure shows (a) a sample test image and predicted (b) GG0, (c) Consolidation, (d) Lung field with the corresponding dice coefficients in parenthesis. Under-segmented regions are shown in red and over-segmented regions in green (e) shows the segmented lobes originally not annotated in the dataset.

(a) CT image slice (b) GT for infection (c) Segmented infection (d) GT for lobes (e) Segmented lobes

Fig. 4. Figure shows (a) a sample slice (s) from Dset5 with annotations for lobes (d), infection (b). In (c), our model's segmented infection and lung region are shown in red and green, respectively, (e) shows the segmented lobes.

- **BL2:** SUMNet[11] trained using a masked segmentation loss.
- **BL3:** 3D UNet trained for segmentation of lung tumor using MSD[4] and segmentaion of COVID infection using COVID-19-CT-Seg dataset[5].
- **BL4:**3D UNet trained for segmentation of lung cancer using StructSeg[6] and segmentaion of COVID infection using COVID-19-CT-Seg dataset.
- **BL5:** 3D UNet trained for segmentation of lungs and pleural effusion using NLCSC[7] and segmentaion of COVID infection using COVID-19-CT-Seg dataset.

Evaluation of **BL3-5** is performed on MosMed (Dset4), for COVID infection segmentation.

C. Pre-processing

The Hounsfield intensity for all scans are clipped between $[-1000, 1000]$ and normalised to $[0, 1]$. The scans in Dset3 of size 630×630 are resized to 512×512. Similar pre-processing was done for Dset1, along with contrast stretching to reduce visual dissimilarity between datasets. The images in Dset2 were normalised to $[0, 1]$ without additional processing.

D. Training Setup

Training is done using Adam optimizer for 30 epochs, with a learning rate of 1×10^{-4}. The trade-off co-efficient λ is set to 10.

[4]http://medicaldecathlon.com/

[5]https://zenodo.org/record/3757476#.Xpz8OcgzZPY

[6]https://structseg2019.grand-challenge.org/

[7]https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=68551327

E. Post-processing

The model is robust to parenchymal involvement in the case of lung segmentation. However, lobe segmentation occasionally misses regions that have infections. These regions are mapped to the lobe with which it shares the maximum boundary.

V. Results and Discussion

Dice coefficient for infection segmentation on the test split of Dset2 and Dset3 are presented in Table I with visualization of prediction in Fig. 3. Additional loss contributed by the auxiliary classifier improves the quality of segmentation, as is evident from the Dice coefficient.

TABLE I

Dice coefficient for segmentation of COVID infections

Method	GGO	Consolidation	Lungs
BL2	43.87	64.26	97.25
Proposed	50.57	64.6	97.71

Performance evaluation of segmentation of COVID infection on Dset4 is shown in Table II. The scores of BL3-5 are

TABLE II

Performance evaluation of segmentation on MosMed

Method	Dice Coefficient
BL2	57.74
BL3	39.2 ± 30.6
BL4	44.3 ± 25.3
BL5	30.1 ± 26.7
Proposed	61.09

TABLE III

PERFORMANCE EVALUATION OF LOBE SEGMENTATION

Method	RU	RM	RL	LU	LL
BL1	92.53	80.6	93.05	96.1	95.3
BL2	84.73	81.73	88.81	83.67	86.31
Proposed	82.6	82.05	85.75	82.22	82.75

TABLE IV

PERFORMANCE EVALUATION ON NEWLY ANNOTATED DATASET.

Method	RU	RM	RL	LU	LL
Proposed	4.38 ± 3.11	1.63 ± 0.82	7.56 ± 7.28	3.26 ± 2.06	5.702 ± 3.11

as reported by the authors[8].

Result of segmentation of the lobes listed in Sec. II are detailed in Table. III. Though the score is lower than BL1, it is to be noted that BL1 was trained solely for lobe segmentation on a single dataset. Our model is trained for lobes and infection segmentation using different data sources.

We annotated lobes in eight volumes (*Dset5*) from COVID-19-CT-Seg dataset with COVID infection annotation. Preliminary annotations obtained using U-net(LTRCLobes_R231) [12] trained on the LTRC dataset were used by a physician to mark the lobes. These annotations, after verification by a radiologist with 15 years of experience, have been made available for download[9]. Qualitative results on Dset5 are shown in Fig. 4. Error e_i in lobe-wise quantification of infection by our proposed model for Dset5 is shown in Table IV.

$$e_i = |G_i - S_i| \tag{3}$$

where G_i is the percentage of infection in lobe i of the ground truth annotation, and S_i is that of the segmented output.

VI. CONCLUSION

We have developed a framework for quantification of parenchymal lesions present in CT scans of COVID-19 patients to provide an objective and rapid score, which is helpful in diagnosis, prognostication and assessment of treatment response. We have achieved this by training a CNN with partial knowledge accumulated from multiple datasets, each having partial annotations needed for the task at hand. In comparison with the prior art, where a network was trained to perform only lobe segmentation, our proposed multi-class segmentation framework segments lobes with an average dice coefficient of 85.68. It segments infections in an unseen dataset with a score of 61.09. We provide lobe annotations for 8 CT volumes to facilitate lobe-wise quantification of infection.

REFERENCES

[1] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," *Radiology*, p. 200642, 2020.

[2] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R.L. Hesketh, L. Yang, et al., "Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (covid-19) pneumonia," *Radiology*, 2020.

[3] H. Tang, C. Zhang, and X. Xie, "Automatic pulmonary lobe segmentation using deep learning," in *IEEE Int. Symp. Biomed. Imag.*, 2019, pp. 1225–1228.

[4] W. Xie, C. Jacobs, J.P. Charbonnier, and B. van Ginneken, "Relational modeling for robust and efficient pulmonary lobe segmentation in ct scans," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2664–2675, 2020.

[5] J. Ma, Y. Wang, X. An, C. Ge, Z. Yu, J. Chen, Q. Zhu, G. Dong, J. He, Z. He, et al., "Towards efficient covid-19 ct annotation: A benchmark for lung and infection segmentation," *arXiv preprint arXiv:2004.12537*, 2020.

[6] N. Lessmann, C.I. Sánchez, L. Beenen, L.H. Boulogne, M. Brink, E. Calli, J.P. Charbonnier, T. Dofferhoff, W.M. van Everdingen, P.K. Gerke, et al., "Automated assessment of co-rads and chest ct severity scores in patients with suspected covid-19 using artificial intelligence," *Radiology*, p. 202439, 2020.

[7] Kai Gao, Jianpo Su, Zhongbiao Jiang, Ling-Li Zeng, Zhichao Feng, Hui Shen, Pengfei Rong, Xin Xu, Jian Qin, Yuexiang Yang, et al., "Dual-branch combination network (dcn): Towards accurate diagnosis and lesion segmentation of covid-19 using ct images," *Med. Image Analysis*, vol. 67, pp. 101836, 2021.

[8] Guotai Wang, Xinglong Liu, Chaoping Li, Zhiyong Xu, Jiugen Ruan, Haifeng Zhu, Tao Meng, Kang Li, Ning Huang, and Shaoting Zhang, "A noise-robust framework for automatic segmentation of covid-19 pneumonia lesions from ct images," *Trans. Med. Imag.*, vol. 39, no. 8, pp. 2653–2663, 2020.

[9] Tanvir Mahmud, Md Jahin Alam, Sakib Chowdhury, Shams Nafisa Ali, Md Maisoon Rahman, Shaikh Anowarul Fattah, and Mohammad Saquib, "Covtanet: A hybrid tri-level attention-based network for lesion segmentation, diagnosis, and severity prediction of covid-19 chest ct scans," *Trans. Industrial Inform.*, vol. 17, no. 9, pp. 6489–6498, 2020.

[10] O. Saha, R. Sathish, and D. Sheet, "Learning with multitask adversaries using weakly labelled data for semantic segmentation in retinal images," in *Int. Conf. Medical Imag. Deep Learn.*, 2019, pp. 414–426.

[11] S. Nandamuri, D. China, P. Mitra, and D. Sheet, "Sumnet: Fully convolutional model for fast segmentation of anatomical structures in ultrasound volumes," in *IEEE Int. Symp. Biomed. Imag.*, 2019, pp. 1729–1732.

[12] H.r Johannes, P. Jeanny, R. Sebastian, P. Helmut, and L. Georg, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *Eur. Radiol. Experimental*, vol. 4, no. 1, 2020.

[8] https://gitee.com/junma11/COVID-19-CT-Seg-Benchmark

[9] http://dx.doi.org/10.21227/3qe9-e178