The Journal of Open Source Software

COVID-19 Lung Segmentation

Riccardo Biondi[*1], **Nico Curti**[†2], **Enrico Giampieri**[2], **and Gastone Castellani**[1]

1 Department of Experimental, Diagnostic and Specialty Medicine of Bologna University **2** eDIMESLab, Department of Experimental, Diagnostic and Specialty Medicine of Bologna University

DOI: 10.21105/joss.03447

Software

- Review
- Repository
- Archive

Editor: Jacob Schreiber

Reviewers:

- @pps121
- @amtseng

Submitted: 16 June 2021
Published: 29 September 2021

Summary

The `COVID-19 Lung Segmentation` project provides a novel, unsupervised and fully automated pipeline for the semantic segmentation of ground-glass opacity (GGO) areas in chest Computer Tomography (CT) scans of patients affected by COVID-19. In the project we provide a series of scripts and functions for the automated segmentation of lungs 3D areas, segmentation of GGO areas, and estimation of radiomic features.

Both PowerShell and bash scripts are provided for the scripts management. A possible Snakemake pipeline for the whole segmentation procedure applied to several CT scans (in a multiprocessing environment) is included into the project.

A detailed description of the whole pipeline of processing has been already discussed in (Biondi et al., 2021@), in which we have showed also the results obtained on public datasets (Jun et al., 2020@). In that work we proved the efficiency of the proposed unsupervised method for the identification of GGO areas and extraction of informative radiomic features. Radiomic features were collected and used to predict clinically relevant scores, with particular focus on mortality and the PREDI-CO score (Bartoletti et al., 2020@).

Statement of Need

COronaVirus Disease (COVID-19) has widely spread all over the world since the beginning of 2020. It is an acute, highly contagious, viral infection mainly involving the respiratory system. Chest CT scans of patients affected by this condition have shown peculiar patterns of Ground Glass Opacities (GGO) and Consolidation (CS) related to the severity and the stage of the disease.

The correct and fast identification of these patterns is a fundamental task. Up to now, this task is mainly performed using manual or semi-automatic techniques, which are time-consuming (hours or days) and subjected to the operator expertise.

This project provides an automated pipeline for the segmentation of GGO areas on chest CT scans of patient affected by COVID-19. The segmentation is achieved with a color quantization algorithm, based on k-means clustering, which groups the voxels by color and texture similarity. This approach is preceeded by the lung segmentation, achieved by a public available U-Net model (Hofmanninger et al., 2020@) (Johannes Hofmanninger, 2020@)

The pipeline performances has been tested on 15 labeled chest CT scans dataset. These scans were segmented and validated by expert radiologist. Ten of these scans were extracted

*co-first author
†co-first author

from the public dataset *COVID-19 CT Lung and Infection Segmentation Dataset* (Jun et al., 2020@) published on Zenodo. Department of Diagnostic and Preventive Medicine of the IRCCS Policlinic Sant'Orsola-Malpighi provided others 82 scans, with the 5 labeled used for the evaluation.

We tested the segmentation performances using the dice coefficient and specificity, sensitivity and precision scores. The average value and the corresponding standard deviation at 1σ are reported in the following Table.

Dice Score	Sensitivity	Specificity	Precision
0.67 ± 0.12	0.66 ± 0.15	0.9992 ± 0.0005	0.75 ± 0.20

The proposed unsupervised segmentation pipeline is able to approximate the gold standard with satisfactory results. Given that the amount of information required for the k-means method training is considerably lower than CNN methods, while still retaining good results, this segmentation can be implemented with in-patient training(Biondi et al., 2021@); as a reference, a 3D U-Net-based method (Yan et al., 2020@) required two order of magnitude training samples to achieve comparable results. With this work we aimed to prove that semi-supervised approaches to segmentation are promising, as they would combine the best effort of highly trained physicians to develop true gold standard segmentation and the expertise of data analysts to augment those segmentation in full blown models. While the proposed pipeline is not yet at the accuracy level necessary for assisted diagnostic, we surmise that our pipeline can be successfully used as starting point for more specific and performing segmentation methods.

Acknowledgments

The authors acknowledge all the members of the Department of Radiology, IRCCS Azienda Ospedaliero-Universitaria di Bologna and the SIRM foundation, Italian Society of Medical and Interventional Radiology for the support in the development of the project and analysis of the data.

References

Bartoletti, M., Giannella, M., Scudeller, L., Tedeschi, S., Rinaldi, M., Bussini, L., Fornaro, G., Pascale, R., Pancaldi, L., Pasquini, Z., Trapani, F., Badia, L., Campoli, C., Tadolini, M., Attard, L., Puoti, M., Merli, M., Mussini, C., Menozzi, M., … group, P. study. (2020). Development and validation of a prediction model for severe respiratory failure in hospitalized patients with SARS-CoV-2 infection: A multicentre cohort study (PREDI-CO study). *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *26*(11), 1545–1553. https://doi.org/10.1016/j.cmi.2020.08.003

Biondi, R., Curti, N., Coppola, F., Giampieri, E., Vara, G., Bartoletti, M., Cattabriga, A., Cocozza, M. A., Ciccarese, F., De Benedittis, C., Cercenelli, L., Bortolani, B., Marcelli, E., Pierotti, L., Strigari, L., Viale, P., Golfieri, R., & Castellani, G. (2021). Classification performance for COVID patient prognosis from automatic AI segmentation—a single-center study. *Applied Sciences*, *11*(12). https://doi.org/10.3390/app11125438

Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., & Langs, G. (2020). Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *European Radiology Experimental*, *4*(1), 50–50. https://doi.org/10.1186/s41747-020-00173-2

Johannes Hofmanninger, H. N. (2020). *Automated lung segmentation in CT under presence of severe pathologies*. https://github.com/JoHof/lungmask; GitHub.

Jun, M., Cheng, G., Yixin, W., Xingle, A., Jiantao, G., Ziqi, Y., Minqing, Z., Xin, L., Xueyuan, D., Shucheng, C., Hao, W., Sen, M., Xiaoyu, Y., Ziwei, N., Chen, L., Lu, T., Yuntao, Z., Qiongjie, Z., Guoqiang, D., & Jian, H. (2020). *COVID-19 CT lung and infection segmentation dataset* (Verson 1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3757476

Yan, Q., Wang, B., Gong, D., Luo, C., Zhao, W., Shen, J., Shi, Q., Jin, S., Zhang, L., & You, Z. (2020). *COVID-19 chest CT image segmentation – a deep convolutional neural network solution*. http://arxiv.org/abs/2004.10987