Segmentation of COVID-19 Lesions in CT Images

Joana Rocha*
INESC-TEC
Faculty of Engineering, University of Porto
Porto, Portugal
0000-0002-4856-138X

Sofia Pereira*
INESC-TEC
Faculty of Engineering, University of Porto
Porto, Portugal
0000-0001-6754-6495

Aurélio Campilho
INESC-TEC
Faculty of Engineering, University of Porto
Porto, Portugal
0000-0002-5317-6275

Ana Maria Mendonça
INESC-TEC
Faculty of Engineering, University of Porto
Porto, Portugal
0000-0002-4319-738X

***The first two authors contributed equally to this work.**

***Abstract*—The worldwide pandemic caused by the new coronavirus (COVID-19) has encouraged the development of multiple computer-aided diagnosis systems to automate daily clinical tasks, such as abnormality detection and classification. Among these tasks, the segmentation of COVID lesions is of high interest to the scientific community, enabling further lesion characterization. Automating the segmentation process can be a useful strategy to provide a fast and accurate second opinion to the physicians, and thus increase the reliability of the diagnosis and disease stratification. The current work explores a CNN-based approach to segment multiple COVID lesions. It includes the implementation of a U-Net structure with a ResNet34 encoder able to deal with the highly imbalanced nature of the problem, as well as the great variability of the COVID lesions, namely in terms of size, shape, and quantity. This approach yields a Dice score of 64.1%, when evaluated on the publicly available COVID-19-20 Lung CT Lesion Segmentation GrandChallenge data set.**

***Index Terms*—computer-aided diagnosis, corona virus, deep neural network, medical imaging, radiology, thorax.**

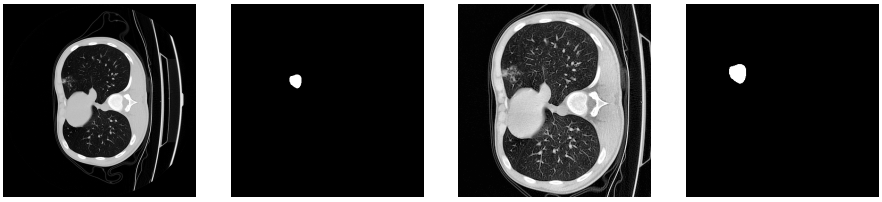
I. Introduction

The new coronavirus disease (COVID-19) has brought devastating worldwide health consequences. As the pandemic progressed, the scientific community quickly started to work on several aspects of the situation, including disease diagnosis and stratification. Considering that the manifestation of the virus in the lungs is among the first signs of infection by SARS-CoV-2 [1], detecting lung abnormalities early on is crucial for clinical management and prognosis, which is why it is a standard procedure for patients to undergo a Computed Tomography (CT) scan. In fact, the CT is currently the gold standard imaging technique to assess lung morphology and detect several associated pathologies, and so it is commonly used to validate emerging automated image analysis techniques [2]. The 3D data resulting from a scan can be analyzed as a whole or alternatively via cross-sectional images (known as slices) retrieved from the volume. Particularly in COVID positive patients, the most common findings in CT scans are ground glass opacities, even though the type and size of the lesions are highly variable and might be correlated with several factors, such as the severity of the infection or the patient's age [1].

Computer-Aided Diagnosis (CAD) systems haven proven to be highly valuable within a diverse range of clinical applications in the past, and therefore it is not surprising that these systems are now being developed for COVID-19 screening, with a strong focus on the analysis of medical image data through deep learning techniques. Image-based CAD systems often include a segmentation task, which seeks to outline a Region of Interest (ROI) represented in the image. Here, the end goal is to obtain segmentation masks for COVID-related lesions in the images. This will later on allow further characterization of the lesions, which can be crucial for evaluating disease severity.

The current benchmark for image analysis are the convolutional neural networks, and when it comes particularly to medical image segmentation, the U-Net is one of the most frequently employed architectures. Its U-shaped architecture has the ability to extract relevant features from the input data and precisely locate the ROI [3]. It is quite challenging to navigate through the work developed in the context of CAD systems for COVID-19 screening. First, the urgency of the matter resulted in a boom of research papers entering the academic literature, some of which are poorly reported and lack proper peer-reviewing. Additionally, most studies developed so far use relatively small data sets, most likely due to annotation and privacy constraints. The lack of statistically relevant data and prior knowledge on the disease puts the developed algorithms and their performance scores at high risk of bias and excessive optimism [4]. Nevertheless, several publications report the use of deep learning approaches for COVID-19 lesion segmentation. Among them, the use of 3D methods based on a V-Net architecture is a possible approach [5]. For instance, one of the most relevant articles at the present moment develops an architecture called VB-Net, which is an adaption of the conventional V-Net [6]. It has gained attention due to its faster execution time, achieved by its integrated bottle-neck structure. Alternatively, opting for 2D approaches is also plausible for

2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-0358-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/BHI50953.2021.9508520



(a) Slice after truncation and normalization.

(b) Slice at the end of the preprocessing.

Fig. 1: Preprocessing stage result for a single CT slice. The HUs were truncated and the pixel intensities normalized (a). Then, the image was center cropped, resized back to 512×512 pixels, and enhanced using the CLAHE technique (b).

reduced computational requirements [7], as exemplified in [8] through the implementation of DeepLabv3. More recently, the COVID-19 Lung CT Lesion Segmentation GrandChallenge achieved very interesting results [9] for this particular exercise that will be used as reference in the current work, whose main contribution lies in the implementation of a 2D U-Net, as a means to reduce the required computational power for COVID lesion segmentation while maintaining a satisfying model performance. A detailed overview of the data set and framework is presented in Section II. The results are further discussed in Section III, and the main conclusions presented in Section IV.

II. Methodology

The data set employed in this work was provided in the COVID-19-20 Lung CT Lesion Segmentation GrandChallenge, whose annotation process resulted from the joint work of the Children's National Hospital (Washington DC, USA), NVIDIA and National Institutes of Health (NIH). The data set contains unenhanced chest CTs from a total of 199 patients with positive RT-PCR for SARS-CoV-2, and ground truth 3D annotations of COVID-19 lesions in the lungs [10].

A. Data Preprocessing

To improve the overall performance of the classifier, several preprocessing steps enumerated in this section were defined and implemented. Please note that the process will be described for a specific patient, but was equally applied to the remaining subjects. First, considering the axial plane, each CT volume consists of multiple slices represented by 512×512 pixels and matching a single patient. It is important to highlight that the current work focuses on a slice-by-slice approach, rather than a straightforward 3D analysis of the data. For this reason, each loaded grayscale image and corresponding ground truth binary mask were decomposed into multiple 2D slices. Additionally, the slices with corresponding null masks were not regarded as input for further lesion delimitation, and were discarded from this data set. The remaining images' Hounsfield Units (HUs) were truncated, meaning that only the intensities between -1000 and 400 were considered. By doing so, foreign bodies (such as gold, steel, and copper) are eliminated from the image, and thus it is easier to analyze the lung structures. Besides truncating the HUs, the intensity values were normalized to the [0,1] range. Considering that the abnormalities are necessarily within the lung and all images are centered, this ROI was further delimited by eliminating all pixels in the outer frame that do not encompass these organs and do not carry relevant information. After cropping the outer pixels, the images were resized back to 512×512 pixels. The same operation was applied to the corresponding masks. Furthermore, the images were enhanced using the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique to ease the identification of the lesions. The CLAHE technique is an adaption of the standard histogram equalization technique that improves the images' contrast while simultaneously avoiding noise amplification. Following the preprocessing stage, the data set is composed of multiple enhanced 2D images and masks per patient (Figure 1), totalling 4981 observations. Furthermore, the four preprocessed masks displayed in Figure 2 draw attention to the high variability in terms of shape, size, and quantity of the lesions present in this data set.

Fig. 2: Example of preprocessed masks, exhibiting the high variability in terms of shape, size, and quantity of the lesions.

B. Experimental Setup

The patients were split into training, validation and test sets to ensure a trustworthy model evaluation, in a way that images belonging to a single patient cannot be present in multiple sets. As such, the test set contains preprocessed images corresponding to 30% of the patients, while the remaining data was used to train and validate the model. More specifically, 5-fold cross validation was used for hyperparameter optimization using the training subjects. These percentage splits ensure that both the validation and test sets have a wide representation of all types of lesions that, as described in this section, present great variability in terms of size, shape, and number. The images in the training set were augmented to increase the number of samples and prevent overfitting, by introducing shear and zoom transformations, as well as vertical and horizontal flips.

TABLE I: Evaluation scores.

Publication	Dice	Jaccard
COVID-19-20 GrandChallenge*	66.6 ± 23.9	54.0 ± 23.6
Proposed U-Net	64.1 ± 26.4	52.1 ± 25.7

*https://covid-segmentation.grand-challenge.org/evaluation/challenge-second-phase-new-data/leaderboard/

The segmentation was performed using a U-Net architecture, since it is considered a state-of-the-art method designed for biomedical image segmentation [3]. The encoder portion of the network was built with a ResNet34 pre-trained on *ImageNet* weights. The data set is highly imbalanced, with approximately 99% of the total number of pixels belonging to the negative class rather than to the class to be segmented, and therefore the network was trained with a Dice-based loss. The optimization of this objective function was performed using the Adam optimizer with an initial learning rate of 10^{-4}. In addition to setting a batch size of 8 and a maximum number of epochs of 100, several callbacks were considered to enhance the model's performance - these include decreasing the learning rate on plateau and interrupting the training routine when the validation loss shows no further improvements. The predictions were obtained using a final sigmoid activation function, establishing a 50% probability threshold for the final pixel classification.

III. Results and Discussion

In order to assess the results, it is necessary to define a specific evaluation system, and consequently decide on the metrics that will be used for such purpose. In the current work, the accuracy, precision, and recall, as well as the Dice coefficient and Jaccard index were selected to further analyze the ability of the network to make correct predictions in a class imbalanced setting. Before presenting the results, it is important to analyze the performance of the model during its training routine, described in Section II. The loss exhibited a convergent behavior and the training process ceased after 23 epochs. In addition, the model did not present any clear signs of overfitting (28% training loss vs. 41% validation loss), and yielded an average Dice score of 61.3% across all validation folds.

The implemented algorithm reported a final test accuracy of 98.7%, recall of 62.4%, and precision of 80.6%. Looking into the individual classes, the proposed methodology presented an accuracy of 99.6% and 62.4% for the majority (background) and minority classes, respectively. The results for the Dice and Jaccard coefficients are displayed in Table I, and are similar to the final scores in the COVID-19-20 Lung CT Lesion Segmentation GrandChallenge leader board. While these results cannot be directly compared, considering that their methodology employed all available 3D data, it is still valuable and relevant to present their values not only to understand the achieved scores, but also the overall difficulty of this particular segmentation exercise.

To illustrate the method's performance, several examples were selected and displayed in this section. Figure 3a-d shows four slices from different patients in which the proposed methodology exhibited a very satisfying performance, while Figure 3e-h shows other cases in which the methodology underperformed. The plots displayed in the mentioned figures highlight the true positive pixels in green, false positive in yellow, and false negative in red. This way, it is possible to understand the differences between the achieved segmentation and the ground truth masks.

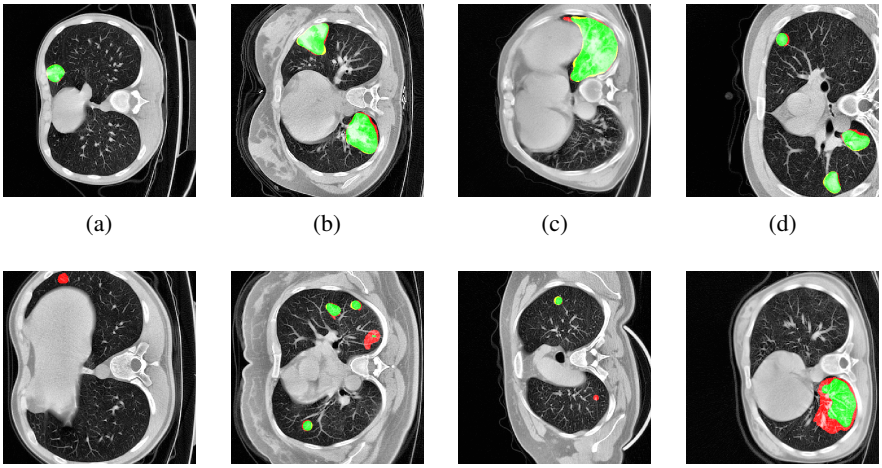
Minding these results, one can infer that the model is able to generalize and correctly predict positive pixels in spite of the lesion's size, shape, or quantity. When the model fails, it misses a certain lesion (Figure 3e-g), or does not recognize its full extent (Figure 3h). Even so, the results proved to be highly consistent, as consecutive slices belonging to the same patient exhibit similar segmentation masks (Figure 4).

IV. Conclusions

The successful segmentation of lung lesions in CT scans of COVID-19 positive cases can be extremely useful for patient diagnosis, prognosis and management. Moreover, being able to do so in an automated fashion can bring crucial advantages in a pandemic context, in which the medical and financial resources are often scarce. The current work aimed to contribute to the location and characterization of these lesions through automated deep learning based segmentation, thus helping the specialists in this highly relevant clinical task.

Consequently, the developed architecture faces three major challenges: the inherent complexity associated with the anatomical structures of the lung, the often noisy, blurred and/or low contrasted nature of medical images, and the significant class imbalance of the data set. To tackle these difficulties, a slice-by-slice 2D approach is considered by implementing a U-Net architecture with transfer learning (i.e. with a pre-trained ResNet34 encoder). Overall, the network is able to correctly segment abnormalities with diverse sizes and shapes, and can easily be incorporated in a CAD workflow following a detection step that removes the null slices. The proposed approach is beneficial mainly in terms of required computational power and more extensively documented in the literature.

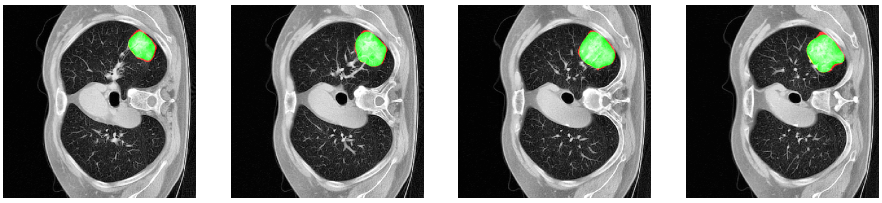
On the other hand, this approach may lead to loss of inter-slice information, and ultimately to a sub-optimal performance. The 2D methodology contributes to the great variability in the lesion types, considering that different slices of a lesion imply smaller and larger representations of the same abnormality. This is due to the fact that the CT slices that do not contain any lesions are excluded from the framework, which also impairs the reconstruction of the result back into a 3D volume. For this reason, the network occasionally miscalculates the lesion's extent or misses smaller lesions. Therefore, future work should focus on the integration of a slice-wise classification step, prior to segmentation, capable of automatically detecting normal slices and excluding them from the lesion segmentation step, but save them for the



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 3: Example of four slices in which the U-Net performed highly (a-d), and performed less satisfactorily (e-h). True positive pixels in green, false positive in yellow, and false negative in red.



(a) (b) (c) (d)

Fig. 4: Example of patient-wise coherence in the segmentation results. True positive pixels in green, false positive in yellow, and false negative in red.

posterior reconstruction of the result into a volume. This would approximate the approach to a 3D scenario without having to change the model's architecture or use additional significant computational resources.

ACKNOWLEDGMENTS

This work was funded by the ERDF - European Regional Development Fund, through the Programa Operacional Regional do Norte (NORTE 2020) and by National Funds through the FCT - Portuguese Foundation for Science and Technology, I.P. within the scope of *UIDB/50014/2020*. The work of J. Rocha was supported by the FCT grant contract *2020.06595.BD*. The work of S. Pereira was supported by the FCT grant contract *2020.10169.BD*.

REFERENCES

[1] Y. Wang, C. Dong, Y. Hu, C. Li, Q. Ren, X. Zhang, H. Shi, and M. Zhou, "Temporal changes of ct findings in 90 patients with covid-19 pneumonia: a longitudinal study," *Radiology*, vol. 296, no. 2, pp. E55–E64, 2020.

[2] G. Bellani, T. Mauri, and A. Pesenti, "Imaging in acute lung injury and acute respiratory distress syndrome," *Current Opinion in Critical Care*, vol. 18, pp. 29–34, 2 2012.

[3] 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*. Springer, 2015, pp. 234–241.

[4] L. Wynants, B. Van Calster, M. M. Bonten, G. S. Collins, T. P. Debray, M. De Vos, M. C. Haller, G. Heinze, K. G. Moons, R. D. Riley, E. Schuit, L. J. Smits, K. I. Snell, E. W. Steyerberg, C. Wallisch, and M. Van Smeden, "Prediction models for diagnosis and prognosis of covid-19 infection: Systematic review and critical appraisal," *The BMJ*, vol. 369, apr 2020.

[5] H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, H. Jiang, D. Wu, H. Sui, C. Zhang, and D. Shen, "Diagnosis of Coronavirus Disease 2019 (COVID-19) with Structured Latent Multi-View Representation Learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2606–2614, aug 2020.

[6] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, "Abnormal lung quantification in chest ct images of covid-19 patients with deep learning and its application to severity prediction," *Medical physics*, 2020.

[7] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, "3D Deep Learning on Medical Images: A Review," *Sensors*, vol. 20, no. 18, p. 5097, sep 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/18/5097

[8] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, Y. Zha, W. Liang, C. Wang, K. Wang *et al.*, "Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography," *Cell*, vol. 181, no. 6, pp. 1423–1433, 2020.

[9] "COVID-19 LUNG CT LESION SEGMENTATION CHALLENGE - 2020 - Grand Challenge," 2020. [Online]. Available: https://covid-segmentation.grand-challenge.org/

[10] P. An, S. Xu, S. A. Harmon, E. B. Turkbey, T. H. Sanford, A. Amalou, M. Kassin, N. Varble, M. Blain, V. Anderson, F. Patella, G. Carrafiello, B. T. Turkbey, and B. J. Wood, "Ct images in covid-19," 2020. [Online]. Available: https://wiki.cancerimagingarchive.net/x/o5QvB