Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
CoLe-CNN+: Context Learning - Convolutional Neural Network for COVID-19-Ground-Glass-Opacities Detection and Segmentation

Giuseppe Pezzano\textsuperscript{a,b,}\textsuperscript{*}, Oliver Díaz,\textsuperscript{b} Vicent Ribas Ripoll\textsuperscript{a}, Petia Radeva\textsuperscript{b,c}

\textsuperscript{a} Eurecat, Centre Tecnològic de Catalunya, eHealth Unit, Barcelona, Spain
\textsuperscript{b} Universitat de Barcelona, Department of Mathematics and Computer Science, Barcelona, Spain
\textsuperscript{c} Computer Vision Center, Bellaterra (Barcelona), Spain

Abstract

\textit{Background and Objective:} The most common tool for population-wide COVID-19 identification is the Reverse Transcription-Polymerase Chain Reaction test that detects the presence of the virus in the throat (or sputum) in swab samples. This test has a sensitivity between 59\% and 71\%. However, this test does not provide precise information regarding the extension of the pulmonary infection. Moreover, it has been proven that through the reading of a computed tomography (CT) scan, a clinician can provide a more complete perspective of the severity of the disease. Therefore, we propose a comprehensive system for fully-automated COVID-19 detection and lesion segmentation from CT scans, powered by deep learning strategies to support decision-making process for the diagnosis of COVID-19.

\textit{Methods:} In the workflow proposed, the input CT image initially goes through lung delineation, then COVID-19 detection and finally lesion segmentation. The chosen neural network has a U-shaped architecture using a newly introduced Multiple Convolutional Layers structure, that produces a lung segmentation mask within a novel pipeline for direct COVID-19 detection and segmentation. In addition, we propose a customized loss function that guarantees an optimal balance on average between sensitivity and precision.

\textit{Results:} Lungs’ segmentation results show a sensitivity near 99\% and Dice-score of 97\%. No false positives were observed in the detection network after 10 different runs with an average accuracy of 97.1\%. The average accuracy for lesion segmentation was approximately 99\%. Using UNet as a benchmark, we compared our results with several other techniques proposed in the literature, obtaining the largest improvement over the UNet outcomes.

\textit{Conclusions:} The method proposed in this paper outperformed the state-of-the-art methods for COVID-19 lesion segmentation from CT images, and improved by 38.2\% the results for F1-score of UNet. The high accuracy observed in this work opens up a wide range of possible applications of our algorithm in other fields related to medical image segmentation.

Keywords:
COVID-19, SARS-CoV-2, Convolutional Neural Network, Segmentation, Detection
1 Introduction

At the end of 2019, we observed the first signs of the worldwide spread of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which continued throughout 2020 and has been protracted to 2021 with more than 170 million confirmed cases and more than 3.5 million deaths (updated on May 28th 2021). Initially, the World Health Organization (WHO) declared the Coronavirus Disease 2019 (COVID-19) as a public health emergency and recognized it as a pandemic only on March 11th, 2020.

The current most widespread technique for COVID-19 identification is the Reverse Transcription-Polymerase Chain Reaction (RT-PCR) test, which detects the presence of the virus through the throat (or sputum) in swab samples. The PCR test does not provide exact information about the severity of the disease, e.g., the spread of the pulmonary lesion caused by COVID-19 [16].

Moreover, once the virus is eliminated by the immune system, the PCR will likely be negative even if the patient is still affected by pneumonia induced by the original COVID-19 infection. It has been shown that this condition can last for weeks [9] [17]. At this point, it has become necessary to develop robust tools based on medical imaging techniques that can provide clinicians with the clinical information they need for properly assessing the progression of the COVID-19 disease.

Two-dimensional (2D) chest X-ray and three-dimensional (3D) thoracic computed tomography (CT) demonstrated to be able to provide a clear picture of the presence, the spread, and the severity of COVID-19 disease [5]. For example, Zheng et al., [30] accurately described how to detect the most common manifestations of this disease, as depicted in Figure 1. Ai et al., [1] observed that, on more than 1,000 suspected cases, only 59% of the patients resulted positive to the RT-PCR test, whereas 88% of them have shown visible COVID-19 traces through the use of CT images. This study has found that the sensitivity of the diagnosis via CT is 97% on the cases predicted with the swab test and that only 3.5% of the patients showed progression on follow-up chest CT scans after RT-PCR test results turned negative. This demonstrates the importance of CT scans, not only in COVID-19 detection, but also in the follow-up of the disease and for assessing the damages after the treatment.

Similarly, Fang et al., [8] stated that 98% of the 51 patients studied showed compat-
Figure 1: a) A 34-year-old female COVID-19 patient presenting fever with dry cough for 2 days. CT scan shows a slight reticular pattern in the left lower lobe and subpleural area (red frame). b) An 81-year-old female COVID-19 patient presenting fever with cough for 7 days. CT scan shows a reticular pattern superimposed on the background of GGO, resembling the sign of crazy paving stones in the right middle lobe (red frame). Extracted from Zheng et al., [30].

Figure 1: a) A 34-year-old female COVID-19 patient presenting fever with dry cough for 2 days. CT scan shows a slight reticular pattern in the left lower lobe and subpleural area (red frame). b) An 81-year-old female COVID-19 patient presenting fever with cough for 7 days. CT scan shows a reticular pattern superimposed on the background of GGO, resembling the sign of crazy paving stones in the right middle lobe (red frame). Extracted from Zheng et al., [30].

1.1 Aim of this work

Lesion segmentation is not useful to diagnose COVID-19, but it is fundamental for assessing the current status of the illness, its severity, and for future treatment planning. In fact, the area covered by the lesion approximately corresponds to the area where the pulmonary alveoli are not working as normal (Liang et al., [17]). Despite the increase in COVID-19’s detection accuracy through the use of CT images, the reading time necessary to interpret 3D CT volumes and to extract the morphological properties of the lesion can greatly increase the workload of radiologists. However, the use of Artificial Intelligence (AI) tools can help to sensibly reduce the interpretation time, as attested by Hosny et al., [13]. Indeed, manual segmentation represents an extremely time-consuming task. Based on a study involving 10 radiologists, Ma et al., [18] found that, on average, one radiologist needs about 400 ± 45 minutes to accurately delineate the lesion in a CT scan with 250 slices. However, deep learning strategies can reduce this time to less than one minute.

The work presented here proposes an automated method for COVID-19 detection using chest CT images, together with the segmentation of the Ground-Glass Opacities (GGO) (Fig.1) and other solidifications/fibrosis present inside the lungs. Thus, a unique Convolutional Neural Network (CNN)-based workflow is built, where the following distinct steps are included:

- Lung segmentation,
COVID-19 detection, and

• COVID-19 lesion segmentation.

The lungs are firstly segmented from the input CT image to reduce the searching area. Afterward, the detection algorithm is used to analyze the lungs’ area in order to detect the presence of COVID-19. In the case of a positive finding, the CT image is processed by the last network (COVID-19 lesion segmentation) to identify the areas affected by the disease.

This paper is arranged as follows: Section 1.2 shows the current state-of-the-art concerning the most relevant methods of COVID-19 detection and lesion segmentation using CT scans. In the following section 2, the methods of COVID-19 detection and segmentation developed in this work, including the model architecture (2.1), are presented. Then, sections 2.2 and 2.4 describe the experimental methods and the dataset used. Next, we present and discuss the results in sections 3 and 4, respectively. In particular, we present the overall performance of the method in section 4.1 and the comparison of the proposed method with the state-of-the-art in section 4.2. Conclusions and future work are finally presented in sections 5 and 6, respectively.

1.2 Relevant Work

COVID-19 has hit the world with such an unbridled force that completely transformed our lives and directed most of the scientific research to counteract its effect. Therefore, the literature on this topic is rapidly increasing, especially in the field of AI, where automated tools can support healthcare professionals in its diagnosis.

We hereby investigated and included the most relevant works about COVID-19 detection and lesion segmentation algorithms using AI using chest CT images. First, we briefly introduce the work of Harrison et al., [12], who tested the expertise of 7 radiologists in discerning between cases of COVID-19 and pneumonia. They did a blinded test using more than 400 CT images, in which the ground-truth was set by RT-PCR test results. The authors gave also an accurate list of differences between COVID-19 and non-COVID-19 viral pneumonia appearance in chest CTs. The results are presented in section 3.

1.2.1 COVID-19 detection

With the term COVID-19 classifier, we refer to one software that classifies the patients in two main classes: positive or neg-
ative to COVID-19, depending on the presence or the absence of the disease. Shi et al., [22] carried out one of the first review articles which suggests the use of neural networks for COVID-19 classification purposes.

Among all analyzed works, we highlight the work of Zheng et al., [32] that developed a two-stage neural network. The first stage for lung segmentation, using UNet architecture [21], and the second one for COVID-19 classification, where only the lung area was considered as the input to the system. For this final step, they provided a mixture between residual blocks and standard convolutions with adaptive max-pooling layers. They trained and tested their algorithm by using a private dataset with remarkable results in terms of accuracy and sensitivity, but low specificity.

Wang et al., [26] adopted a modified Inception network [24] [25] to classify the patients into COVID-19 or pneumonia classes. They also made use of a private CT image dataset of nearly 100 patients, in order to fine-tune and then test the network. They compared the predictions of their neural network with those provided by two radiologists, on a total of 745 images. Moreover, they reported a study according to which 75% of patients with negative RT-PCR results were positive to COVID-19 from findings in their CT images.

Wu et al., [28] provided an explicable neural network, based on VGG-16 [23] that classifies the patients into COVID-19 and non-COVID-19, and returns a heat map to highlight the areas, where the lesion is present. With the introduction of a large dataset (810 images) and under deep supervision, they registered a maximum of 96% for sensitivity, but specificity was below 94% for all the cases. They additionally provided a COVID-19 segmentation algorithm that will be presented in section 1.2.2. Hu et al., [14] used a 5-layers convolutional neural network for infection detection and classification. The encoded information was extracted from the last layer of the three last levels of depth, then this was concatenated and used to predict the patient status. With the aid of a well-constructed data augmentation strategy, they presented their results for classification between COVID-19, no-COVID-19, and pneumonia cases. They also included a weight that is dependent on the class frequency of occurrence, as well as on the loss function.

1.2.2 COVID-19 segmentation

In this work, COVID-19 segmentation refers to the delineation of the lesion induced by the presence of coronavirus into the lungs. As described by Kong and Agarwalhey [15].
the lesion appears in two different shapes: (1) as a GGO and (2) as a solidification of the tissue. Although they are classified independently one from the other, we do not distinguish between them in this work because they both are COVID-19 manifestations.

Compared to our proposed pipeline, the most similar work we found in the literature is from Fan et al., [7]. They used the same open-access image databases and, in the same way, they employed a lung segmentation algorithm as a first step. The InfNet that they proposed is a combination of Reverse Attention modules for edge learning and parallel decoders with deep supervision. In order to provide a more comprehensive picture of their segmentation performance, they also compared their results with the most widespread versions of UNet.

Wu et al., [28] used a large dataset of 810 CT images, which is partially available online. Their network for predicting the lesion masks for COVID-19 consists of a VGG-16 encoder, similar to their classification network, and a 5-levels-of-depth decoder.

The COVID-19 Pneumonia lesion (COPLE-Net) segmentation algorithm, designed by Wang et al., [27], was also inspired by UNet, but with the addition of extra residual connection and a reduced number of filters carried from the encoder. Their method was further enforced with noise-robust features and loss function. They trained and tested their method on a large private dataset.

Chen et al., [3] provided a residual attention UNet, where the convolutions have been substituted with ResNeXt Blocks. They used a relatively small dataset of 110 CT images, but with intensive work on data augmentation. A similar work was conducted by Zhou et al., [33], using the same dataset and the same network structure.

Yan et al., [29] proposed a Feature Variation (FV) block that enhanced the capability of feature representation adapting to diverse cases. They also applied Progressive Atrous Spatial Pyramid Pooling for handling various infection areas with diverse appearances and shapes. Their dataset included more than 800 CT scans from five different hospitals in China, whose manual annotations and segmentation (i.e. ground truth) were performed by six expert radiologists.

Elharrouss et al., [6] used two encoders and one decoder for lung and COVID-19 lesion segmentation. The inputs of the lung segmentation model correspond to the CT slice image and its texture. In addition, the COVID-19 lesion segmentation model required the CT image and the segmented lungs. They used a subset of a publicly avail-
able dataset of 100 CT images.

The experimental part of Ma et al., [18] is likewise similar to the one proposed here. They first trained their network for lung segmentation (with an average of 86.91% in Dice-score) and then, after filtering out the part of the CT external to the lungs, they performed COVID-19 lesion segmentation. This network represents a revisited version of UNet with minor changes, trained and tested five times on a publicly available dataset that has been prepared by the authors specifically for this task.

Yu et al., [31] proposed a multi-class COVID-19 segmentation network in the shape of a classic encoder-decoder structure, with a pyramid attention mechanism and a loss function based on wavelet decomposition. The pyramid attention module combines a pyramid multi-scaling and channel attention mechanism to highlight salient features at each stage, in addition to a wavelet edge loss function, which uses wavelet decomposition to extract multi-directional edge information of the lesion area to improve the accuracy of the segmentation. They use two datasets: one public, used also in our work, and one private. We report in section 3 the statistics only for the public dataset.

Despite the publication of similar AI strategies in the literature to detect and segment COVID-19 from CT images, we observed a general lack of open access datasets, which would facilitate a fair comparison by members of the community. To train and validate the methods proposed in this work in a robust and reproducible way, we created a multi-centre and multi-vendor CT image dataset from publicly available datasets, as described below.

2 Methods

Before explaining the architecture used in this work, we want to underline that all the input images of the network are CT axial images normalized in the interval $[0, 1]$ (where a lower value corresponds to a medium with a lower electron density) and re-sampled at a resolution of $256 \times 256$ pixels.

2.1 Model Architecture

In Figure 2, we present the model architecture used for lung and COVID-19 lesion segmentation. It corresponds to a $U$-shaped network with 4 levels of depth. At each level of the encoder, there is a Multiple Convolutional Layer (MCL) and a pooling layer. The MCL, shown in Figure 3, has four parallel branches. In three of these, the input tensor is convoluted from one to three times with a series of $3 \times 3$ convolutions. In the fourth branch, it is simply copied with an identity function (marked as a red cube). The four
Figure 2: Structure of our CoLe-CNN+, with one encoder and two decoders. The name on each sample corresponds to the operation applied. The last step is always a convolution. The number of filters is provided between brackets. Input and output, convolutional and re-scaling layers are indicated in green, blue and red, respectively. The abbreviations conv, cat and up correspond to convolution, concatenation and up-sampling, respectively.

Figure 3: Structure of our MCL, with four branches. The name on each sample corresponds to the operation applied in order to obtain it. The symbol 1 represents the Identity function repeated the correspondent number of times. The relative number of filters is provided between brackets. With \( n \) we indicate the input number of filters.

Results are then concatenated and convoluted one more time. The number of filters is doubled at each MCL, except for the first layer where the number of filters goes from 1 to 40. After this block, the tensor is pooled with a 2 × 2 kernel and then given to the following level. The last stage is a 3 × 3 convolution, commonly called bridge. Afterward, the tensor passes to the decoders. These decoders are two almost identical parallel branches providing two different outputs. In each decoder, the input is up-sampled and concatenated with the output from the MCL of the correspondent level of depth, then it is convoluted again. This step is repeated four times.
in order to restore the original shape of the image. The only difference between the two decoders lies in the size of the kernels of the convolutions, which is $3 \times 3$ in one branch and $5 \times 5$ in the other one.

The architecture of the segmentation algorithms is inspired by the CNN used in Pezzano et al., [20] and provides some major differences developed specifically for the purposes of COVID-19 lesion segmentation. The key differences from the architecture used in our previous work are: (i) the use of a loss function with a parameter for maximizing sensitivity, (ii) a new architecture of the MCL, (iii) a new mask calculation formula (explained in the final part of section 2.1), (iv) an additional post-processing procedure to reduce false positives and increase specificity, (v) two additional levels of depth of the network, which then go from two to four levels, and (vi) an extensive validation with an accurate selection of hyper-parameters derived specifically for the purposes of COVID-19 lesion segmentation, i.e. learning rate, batch size, thresholds, etc.

In our architecture, each decoder $j$ produces two masks ($M_{j,1}$ and $M_{j,2}$) that are then used to construct the COVID-19 lesion segmentation mask $M_{cov}$ with the following formula:

$$M_{cov} = I \cdot M_{1,1} + M_{2,1} - I \cdot M_{1,2} - M_{2,2} = I(M_{1,1} - M_{1,2}) + M_{2,1} - M_{2,2},$$

where $I$ is the input CT image, and the operators $\cdot$, $-$ and $+$ are the element-wise product, subtraction, and addition, respectively. The mask is then converted into a binary image using a threshold set at $0.1$. In the case of lung segmentation, equation (1) is slightly different, because we want to segment an area whose values are near to zero intensity, since the volume inside the lung, consisting mainly of air, is characterized by a very low density. Thus, we calculate the mask starting from $1$ (a tensor of unitary values only, with the same shape as the input) minus the input image, as shown below:

$$M_{cov} = (1 - I)M_{1,1} + M_{2,1} - (1 - I)M_{1,2} - M_{2,2} = I(M_{1,2} - M_{1,1}) + M_{2,1} - M_{2,2}$$

The structure of the decoder changes for COVID-19 classification purposes (Figure 4), due to the need of using Fully Connected Layers (FCLs). This decoder takes the result of the bridge operation and convolves it three
Figure 4: Structure of the decoder for COVID-19 detection. The name on each sample corresponds to the operation applied in order to obtain that sample. In parentheses, the relative number of filters appears.

2.2 Implementation Details

The lung and COVID-19 segmentation networks have been coupled with a composite loss \( L \) that minimizes the Mean Square Error (MSE) and the sensitivity \( S \):

\[
L = \sum_i \left[ MSE(z_i, \hat{z}_i) + 0.5(1 - S(z_i, \hat{z}_i)) \right],
\]

\( i \) is the index that runs over all the pixels of the images, \( z \) and \( \hat{z} \) are the elements of the predicted mask and the ground truth.

The classification network uses the Cross Entropy (CE) loss, which is defined as follows:

\[
CE(x, class) = -x[\text{class}] + \log \left( \sum_j e^{x[j]} \right),
\]

where the classes are two, namely COVID-19 and no-COVID-19.

At the beginning of each training, the dataset has been randomly divided into training and test sets with a ratio of 1 \( \div \) 8.

Due to the limited number of data, we operated a data augmentation strategy. At the beginning of each epoch, all the CT slices of the training test have been randomly rotated within the interval \([-180^\circ, 180^\circ]\) and randomly zoomed with a magnification included in the range \([\times 0.95, \times 1.05]\). The eventual padding has been made with black pixels (equivalent to zero). As a next step, we operated another transformation using the \texttt{torchvision.transforms.ColorJitter} function. This function randomly changes the brightness, the saturation and the contrast of a factor of \(\pm 0.2\). This helps the network to learn how to work with multi-vendor, multi-centre input data with different exposure times, dose and reconstruction methods.

In our experiment, the total number of learnable weights for COVID-19 detection is 15,427,042 for a total of \(\sim 62MB\) of memory occupation. For segmentation, the number of weights is 5,504,074 for \(\sim 22MB\).
We trained our CNNs during 100 epochs in the classification case and 60 epochs in the segmentation case. In order to support our expensive calculations, we used a machine equipped with an Intel Xeon E7-4830 v3 (i7) processor with 64GB of RAM. The GPUs are two Nvidia GeForce GTX 1080 Ti, each one with 12GB of CUDA memory. All the libraries employed were implemented in Python 3.7 with the usage of Pytorch module. We decided to use a batch size of 2. This allowed us to gain better results at the expenses of the computation time. The learning rate has been fixed at 0.00001 for classification and 0.0005 for both segmentation neural networks. The average time consumption for a single segmentation training was 180 minutes, and 30 minutes for classification. The time spent for testing on a single image (once the model has been loaded\footnote{\begin{small}Loading the model is an operation that can take up to one minute on our hardware and needs to be done only once for every test set.\end{small}}) is less than a second for all the networks. We used Adam optimizer for back propagation, since it ensures the stability in non-stationary problems even in the presence of sparse gradients.

### 2.3 Post-Processing

In order to increase the accuracy of our methods, the resulting segmentation masks were further processed. This step allows us to reduce the image noise, which is present in the form of sparse points with no relevance to the actual segmentation. Using the morphological transformations module of OpenCV (v. 4.3.0), we apply a first softer operation of opening (first erosion and then dilation) with the $3 \times 3$ kernel $K_1$ (Equation (5)). Then, in order to obtain smoother contours and to avoid holes within the segmented area, we operated a stiffer transformation of closing, using the $K_2$ kernel (Equation (5)):

\[
K_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad K_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (5)
\]

### 2.4 Image Dataset

For this study, we created a multi-vendor, multi-centre COVID-19 CT image dataset after combining data from several public databases. Due to the novelty of the research topic, open-access COVID-19 CT image repositories are difficult to find. We obtained access to a total of 79 COVID-19 CT volumes and 110 CT slices that have sufficient annotation for our purposes. Among these, 20 volumes come from Cohen et al., and all of them have COVID-19 lesion and lung segmentation annotations.
we downloaded 1110 CT images but only 50 of them were useful for our purpose, having COVID-19 lesion segmentation. The remaining 9 CT volumes and 110 slices were collected from SIRMI and include lung and lesion masks. In order to have a pool of CT scans of patients not suffering from COVID-19 or pneumonia (healthy cohort), we included in our dataset 884 CT images from LIDC-IDRI dataset, described in Armato et al., [2]. These CTs have also complete lung masks, thus we employed them for both classification and lung segmentation.

Since our model is for 2D image processing, it allows us to take different slices from the same CT volume, but with the constraint that 2D slices belonging to the same CT, i.e. the same patient, cannot be found both in the training and the test set. The datasets have been split at patient level for the cross-validation. In all the cases, the images extracted from one CT, i.e. from one patient, can belong strictly to the train set or to the test set only. There are no repeated CTs in our datasets, neither the intersection of patient’s data in both training and test datasets. In this case, we automatically selected 558 slices from 79 CT volumes, using the criteria that the area of the annotated COVID-19 lesion must be larger than 20 pixels at a resolution of $256 \times 256$, and the distance between the centers of the slices must be larger than 30 mm. In order to get a larger set of images for lung segmentation, we extracted three slices from each CT of the LIDC-IDRI dataset. Therefore, the total number of 2D CT images used in this work is: 663 with COVID-19 and 2,652 without COVID-19. All of the aforesaid slices have been taken on the axial plane.

It has recently become a standard procedure in medical image segmentation to compare the results with those obtained with UNet. Indeed, UNet is a reliable network which often represents the skeleton of more modern architectures. Due to the absence of unified COVID-19 -CT databases as well, an objective method to make a comparison between different works would be to use UNet as a benchmark.

### 2.5 Evaluation Statistics

After the post-processing stage, a binary image is created using a threshold set at 0.5 for lung segmentation and at 0.1 for COVID-
19 lesion segmentation. Such binary image is then compared to the ground truth \( G \) in a pixel-by-pixel manner. The evaluation criterion most commonly used in this type of problems is the sensitivity \( (\text{Sens}) \). High sensitivity is desired, although this could lead to over-segmentation. For this reason, this must be balanced with precision \( (\text{Prec}) \). Their reduced mass is proportional to the \( F1 \) (or Dice Similarity Coefficient - \( \text{DSC} \)). In addition to these evaluation metrics, we calculated the accuracy \( (\text{Acc}) \) and the specificity \( (\text{Spec}) \) of the prediction. In terms of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) predictions\(^8\), we used the following statistics:

\[
\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN},
\]

\[
\text{Spec} = \frac{TN}{TN + FP},
\]

\[
\text{Prec} = \frac{TP}{TP + FP},
\]

\[
\text{Sens} = \frac{TP}{TP + FN},
\]

\[
F1 = \frac{2 \cdot \text{Sens} \cdot \text{Prec}}{\text{Sens} + \text{Prec}},
\]

(6)

Another widespread evaluation criterion is the Intersection over Union (\( \text{IoU} \)) described by:

\[
\text{IoU} = \frac{G_+ \cap M_+}{G_+ \cup M_+},
\]

(7)

where \( G \) is the ground truth, \( M \) is the mask and \( + \) is the index indicating only the pixels where the values of the tensors are equal to 1. Although IoU does not provide more information than F1, it is extensively used since it is also invariant to changes in the entries: \( \text{IoU}(G, M) = \text{IoU}(M, G) \).

### 3 Results

#### 3.1 Lung Segmentation

Since COVID-19 mainly affects the inside part of the lungs, the pleura and all the background have been filtered out from the CT images before training the classification and the segmentation networks. In order to provide a fully automated method, the lung masks have been calculated with a neural network identical to the one used for lesion segmentation. The lung masks have been generated and the our results are shown in Table 1.

| Cases | IoU | Acc | Sens | Prec | F1 |
|-------|-----|-----|------|------|----|
| 392   | 94.7| 99.4| 98.9 | 95.4 | 96.9|

Table 1: Average results of the algorithm used for lung segmentation, over the number of cases. Statistics are given in percentage.
Table 2: Result of the 10 runs of our COVID-19 lesion segmentation algorithm. Statistics are given in percentage.

Table 3: Results of the 10 runs of our COVID-19 detection algorithm. Statistics are given in percentage.

3.2 COVID-19 lesion Segmentation

Table 2 shows the results of ten runs of our COVID-19 lesion segmentation network, measured with the evaluation metrics presented in Section 2.5. For each run, the dataset was randomly divided into training and test sets with a proportion of 1/8. In addition, the slices belonging to the same CT were not used in both training and test set. Then, for each epoch, we randomly operated data augmentation as explained in Section 2.2. Figures 5 and 6 present sample results for lesion segmentation.

3.3 COVID-19 detection

The results after ten runs of training of COVID-19 detection are depicted in Table 3. In this case, the ratio between the number of CT slices with and without COVID-19 is 1:6.

4 Discussion

In this section we present an extensive analysis of our results compared with the most recent related articles in the literature.

4.1 Overall Performance

As previously described, the input of our CoLe-CNN+ was filtered using a prior lung segmentation step. Through the use of a sensitivity parameter in the loss function, in Equation 3, we obtained a sensitivity of almost 99%, in average. This result, coupled with a high accuracy, shows that the loss of information within the inside part of the lungs is negligible in the majority of the cases. This lung segmentation step has two main advantages for the COVID-19 lesion segmentation: (i) an improvement in accu-
racy and specificity (Table 2), because false positives in the background are minimized; and (ii) a reduction of the probability to have under-segmentation of the lesions, e.g. if an incorrect lung mask considers a COVID-19 lesion as part of the pleura and filter it out, that lesion would not be detected. The values of sensitivity and precision are balanced, on average. This means that our algorithm has generally found the optimum equilibrium for a reliable prediction.

Table 3 shows the COVID-19 detection results of ten runs. Notice that, in all the cases, the network produces zero FPs. Setting specificity, precision and Positive Predicted Value (PPV) to a fixed 100% is a fundamental requirement in this kind of tools. Our reported accuracy in COVID-19 detection is higher than 95% in all the cases, with a very small standard deviation stating the robustness of our method. Also, the average sensitivity is 78%. One of the reasons for the difference between accuracy and sensitivity lies in the large disparity in the CT numbers (i.e. pixel intensities) for COVID-19 and no-COVID-19 subjects.

4.2 Comparison to the State of the Art

The definitive proof of reliability of our method is given by the comparison of our results with the current state of the art. In Table 1, we compare our average and the best results to four articles, described in section 1.2.1 and a work from Harrison et al., [12], where ten radiologists detected COVID-19 during CT reading sessions. We can infer that our results outperform the state of the art in almost all the evaluation metrics and ensure a very strong reliability with solid results in precision and PPV. Particular relevance should be given to the accuracy level which is almost 8% better than the maximum value obtained in the other works from the literature. The results of sensitivity are, however, in line with the other works. The highest results in this column have been achieved by Hu et al., [14] and Wu et al., [28]. With regard to this last one, we cannot discuss the general goodness of the work because of the absence of information regarding the other parameters. About the latter, we can attest that it is balanced in sensitivity, specificity and precision and this suggests that it can be considered reliable. The same assertion can be done for Zheng et al., [32]. Despite all, considering the results of the radiologists and all the information available to us, our work has shown remarkable results. As mentioned in section 2.4 we trained and tested a UNet architecture using the same data augmentation that we used for our
Table 4: Average results in classification from radiologists and other major articles in this field. Results are given in percentage. ∗ In Harrison et al., [12], the classification has been made between COVID-19 and pneumonia cases.

| Paper        | CTs | Acc | Sens  | Prec | Spec | PPV | NPV |
|--------------|-----|-----|-------|------|------|-----|-----|
| Harrison et al., ∗ [12] | 424 | 74.3 | 79.3  | -    | 68.7 | 78.7 | 76.3 |
| Wang et al., [26]     | 58  | 82.1 | 80.4  | -    | 83.7 | 89.8 | 78.6 |
| Zheng et al., [32]    | 453 | 73.1 | 67.1  | -    | 76.4 | 61.0 | 81.0 |
| Wu et al., [28]       | 630 | 90.8 | -     | -    | -    | 86.7 | 96.4 |
| Hu et al., [14]       | 810 | -   | 96.0  | -    | 91.5 | -    | -    |
| our average           | 189 | 97.1 | 78.0  | 100  | 100  | 100  | 96.8 |
| our best              | 189 | 98.5 | 87.5  | 100  | 100  | 100  | 98.3 |

method. The comparison between the methods described in Section 1.2.2 and UNet is shown in Table 5. As expected, the results obtained by UNet are very different depending on the dataset used. We can see that the F1-score floats between 40.1% and 82.0%. A difference of almost 42%, using the same method, does not only depend on data augmentation or hyper-parameter setting but also it is influenced by data annotation, ground truth creation and the variety of cases. Our choice to use a multi-vendor and multi-centre image dataset has been made also in consideration of this aspect. Therefore, our purpose is to develop a reliable method that can work with data acquired from several and different sources. In fact, from Figures 5 and 6, an expert eye can notice that, although all the images have been all pre-processed in an identical way, there are significant differences between them in terms of image properties, acquisition methods, resolution and, most importantly, annotation methods. These big differences between the image datasets used can affect the inferences deducted from the bare comparison between the performances of the proposed networks. The direct comparison with the literature tells us that our method is, on average, the most sensible, specific and accurate one. Particular emphasis should be given to the balance between sensitivity and precision, and the level of specificity which ensure the almost complete absence of gross errors of segmentation, especially in areas external to the lungs. This is already a distinctive accomplishment, but the data that really prove the goodness of the proposed method are shown in Table 6. Excluding accuracy and specificity that were already well over 96%, the improvement over the results of UNet is significant in each field and
greater than the ones obtained by all the other methods compared. Indeed, we gain an improvement of 38.2\% in F1-score and 28.8\% in sensitivity. That means 3.7\% and 11.1\% more (in F1-score and sensitivity, respectively), with respect to the best method that we have compared.

The strongest limitation of our work is the quantity and the quality of the datasets used. This not only reduces the number of examples that our network can study but also does not allow us to perform deeper studies, such as the inter-observer variability or the accuracy of our method based on the area of the lung afflicted by COVID-19 lesions. Furthermore, with a unique and publicly available dataset, we could have done a direct and fairer comparison with other works. Another limitation is given by the hardware used for the experimental part. Implementing a similar method directly on 3D CT volumes, could theoretically give more accurate results. However, our available hardware did not allow us to train a network of such complexity on a whole 3D CT image, or at full resolution of $512 \times 512 \times 512$ voxels.
Figure 5: Examples of the results obtained with CoLe-CNN+ (in green) compared with the ground truth (in red). In yellow, the superposition of the two segmentations.
Figure 6: Other examples of the results obtained with CoLe-CNN+ (in green) compared with the ground truth (in red). In yellow, the superposition of the two segmentations.
| Paper          | Method  | CTs | IoU   | Acc    | Spec   | Sens   | Prec   | F1-score |
|---------------|---------|-----|-------|--------|--------|--------|--------|----------|
| Fan et al., [7] | UNet   | 129 | -     | -      | 85.8   | 53.4   | -      | 43.9     |
|               | Semi-Inf-Net | 129 | -     | -      | 96.0   | 72.5   | -      | 73.9     |
| Wu et al., [28] | UNet   | 810 | 54.1  | -      | -      | -      | -      | 65.1     |
|               | JCS     | 810 | 66.5  | -      | -      | -      | -      | 78.3     |
| Wang et al., [27] | UNet  | 558 | -     | -      | -      | -      | -      | 70.3     |
|               | COBLE-Net | 558 | -     | -      | -      | -      | -      | 80.3     |
| Chen et al., [3]† | UNet   | 110 | -     | 83.0   | -      | -      | 79.0   | 82.0     |
|               | theirs  | 110 | -     | 95.0   | -      | -      | 89.0   | 94.0     |
| Yan et al., [29] | UNet   | 861 | -     | -      | -      | 73.6   | 66.2   | 68.8     |
|               | theirs  | 861 | -     | -      | -      | 75.1   | 72.6   | 72.6     |
| Elharrouss et al., [6] | UNet | 98  | -     | -      | 85.8   | 53.4   | -      | 43.9     |
|               | theirs  | 98  | -     | -      | 99.3   | 71.1   | 85.6   | 78.4     |
| Ma et al., [18] | mod UNet | 161 | -     | -      | -      | -      | -      | 67.3     |
| Yu et al., [31] | theirs  | 100 | -     | -      | 98.3   | 79.1   | -      | 77.9     |

Table 5: Average results for our segmentation algorithm compared with the SoA. Results are given in percentage. †Chen et al., [3] calculated F1-score with a different formula that takes into account not only the positively predicted pixels, but also the negative ones, that makes difficult the direct comparison. This explains also the unexpected closeness of F1 with the accuracy.

| Paper          | IoU | Acc | Spec | Sens | Prec | F1-score |
|---------------|-----|-----|------|------|------|----------|
| Fan et al., [7] | -   | -   | +10.2% | +19.1% | -   | +30.0%   |
| Wu et al., [28] | +12.4% | -   | -    | -    | -    | +13.2%   |
| Wang et al., [27] | -   | -   | -    | -    | -    | +10.0%   |
| Chen et al., [3] | -   | +13.0% | -    | -    | +10.0% | +12.0%   |
| Yan et al., [29] | -   | -   | +1.5% | +6.4% | +3.8% | +3.8%    |
| Elharrouss et al., [6]† | -   | -   | +13.5% | +17.7% | -   | +34.5%   |

Table 6: Average increment for each segmentation algorithm compared with UNet, tested on the same dataset.
5 Conclusions

In this work, we propose a novel routine for lung segmentation, COVID-19 detection and lesion segmentation. Thanks to an accurate lung segmentation and a sensitivity near 99%, we have been able to reach a PPV value of 100%. We did not observed any FP on the ten different runs of the network, with an average accuracy of 97.1 ± 1.0%. Our sensitivity was therefore in concordance with the one achieved in average by a pool of 10 radiologists. Regarding lesion segmentation, we proved that our method is, in absolute numbers, competitive with the best current methods in the state of the art, with an average accuracy of 98.9 ± 0.3%. In addition, we showed that our method improved the results obtained by UNet on the same dataset, by the largest quantity. In fact, we gained 38.2% on F1-score and 36.1% on IoU, over UNet.

6 Outlook

The robustness and the accuracy of this work open up a wide range of other possible applications of our method. For example, the proposed network could be adapted, using fine-tuning, for studying the worst cases of pneumonia, the diffused metastasis or other lung diseases. Also, our methodology could be applied to detect and segment a large variety of organs in other fields of medical imaging analysis.

Declaration of Competing Interest

None.

Acknowledgements

This work was financially supported by the Catalan Government through the funding grant ACCIÓ-Eurecat, TIN2018-095232-B-C21, SGR-2017 1742, and the Greenhabit project funded by EIT Digital program. Giuseppe Pezzano is a fellow of Eurecat’s Vincente López PhD grant program. Credits to Mozorov et al., [19], for providing the MosMedData. More info here.

List of Abbreviations

- CT: Computed Tomography
- CNN: Convolutional Neural Network
- MCL: Multiple Convolutional Layer
- CoLe: Context Learning CNN
- IoU: Intersection over Union or Jaccard
- Acc: Accuracy
- Spec: Specificity
- Sens: Sensitivity
- Prec: Precision
- F1: F1-score or Dice score
- PPV: Positive Predicted Values
- NPV: Negative Predicted Values
References

[1] Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L - Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases - https://doi.org/10.1148/radiol.2020200642 [2020]

[2] Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA, MacMahon H, van Beek EJR, Yankelevitz D, et al., - The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans - Physics, 38: 915–931 [2011]

[3] Chen, X., Yao, L., Zhang, Y. - Residual Attention UNet for Automated Multi-Class Segmentation of COVID-19 Chest CT Images - arXiv:2004.05645 [cs, eess, q-bio] [2020]

[4] Cohen, J.P., Morrison, P., Dao, L. - COVID-19 Image Data Collection - arXiv:2003.11597 [cs, eess, q-bio] [2020]

[5] Cuipeing Bao, Xuehuan Liu, Han Zhang, Yiming Li, Jun Liu - Coronavirus Disease 2019 (COVID-19) CT Findings: A Systematic Review and Meta-analysis - Journal of the American College of Radiology, Volume 17, Issue 6, Pages 701-709, [2020]

[6] Elharrouss, O., Subramanian, N., Al-Maadeed, S. - An encoder-decoder-based method for COVID-19 lung infection segmentation - arXiv:2007.00861 [cs, eess] [2020]

[7] Fan D.P., et al., - Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images - arXiv:2004.14133 [2020]

[8] Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., Ji, W. - Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR - Radiology 296, E115–E117 [2020]

[9] Fu, Z., Tang, N., Chen, Y. et al., - CT features of COVID-19 patients with two consecutive negative RT-PCR tests after treatment. - Sci Rep 10, 11548 [2020]

[10] Goodfellow I, Bengio Y, Courville A - Deep Learning - MIT Press [2016]

[11] Han, G., Liu, X., Han, F., Santika, I.N.T., Zhao, Y., Zhao, X., Zhou, C. - The LISS—A Public Database of Common Imaging Signs of Lung Diseases for Computer-Aided Detection and Diagnosis [Research and Medical Education]. - IEEE Transactions on Biomedical Engineering 62, 648–656 [2015]

[12] Harrison X.B., et al., - Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT - arXiv:2003.13865v15 [2020]

[13] Hösny, Ahmed and Parmar, Chintan and Quackenbush, John and Schwartz, Lawrence H and Aerts, Hugo JWL - Artificial intelligence in radiology - Nature Reviews Cancer 18, 500-510 [2018]

[14] Hu, S., Gao, Y., Niu, Z., Jiang, Y., Li, L., Xiao, X., Wang, M., Fang, E.F., Menpes-Smith, W., Xia, J., Ye, H., Yang, G. - Weakly Supervised Deep Learning for COVID-19 Infection Detection and Classification From CT Images - IEEE Access 8, 118869–118883 [2020]

[15] Kong, W., Agarwal, P.P. - Chest Imaging Appearance of COVID-19 Infection - Radiology: Cardiothoracic Imaging 2, e200028 [2020]

[16] Lassau, N., Ammari, S., Chouzenoux, E. et al., - Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. - Nat Commun 12, 634 [2021]

[17] Liang, T., Liu, Z., Wu, C.C. et al., - Evolution of CT findings in patients with mild COVID-19 pneumonia. - Eur Radiol 30, 4865–4873 [2020]

[18] Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z., Nie, Z., Yang, X. - Towards Efficient COVID-19 CT Annotation: A Benchmark for Lung and Infection Segmentation - arXiv:2004.12537 [cs, eess] [2020]
[19] Morozov, S., Andreychenko, A., Blokhin, I., Vladzymyrskyy, A., Gelezhe, P., Gombolevskiy, V., Gonchar, A., Ledikhova, N., Pavlov, N., Chernina, V. - MosMedData: Chest CT Scans with COVID-19 Related Findings - v. 1.0 [2020]

[20] G. Pezzano, V. Ribas Ripoll, P. Radeva - CoLe-CNN: Context-learning convolutional neural network with adaptive loss function for lung nodule segmentation - Computer Methods and Programs in Biomedicine, Volume 198, 105792, ISSN 0169-2607 [2021]

[21] Ronneberger O, Fischer P, Brox T - U-Net: Convolutional Networks for Biomedical Image Segmentation - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham [2015]

[22] Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., Shen, D. - Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for COVID-19 - IEEE Reviews in Biomedical Engineering 1–1 [2020]

[23] Simonyan, K., Zisserman, A. - Very Deep Convolutional Networks for Large-Scale Image Recognition - arXiv:1409.1556 [2015]

[24] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA - Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning - Presented at the Thirty-First AAAI Conference on Artificial Intelligence [2017]

[25] Szegedy C, Ioffe S, Vanhoucke V, Shlens J - Revisiting the Inception Architecture for Computer Vision - CVPR [2016]

[26] Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Cai, M., Yang, J., Li, Y., Meng, X., Xu, B. - A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19) - Infectious Diseases [2020]

[27] Wang, G., Liu, X., Li, C., Xu, Z., Ruan, J., Zhu, H., Meng, T., Li, K., Huang, N., Zhang, S. - Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions From CT Images - IEEE Transactions on Medical Imaging 39, 2653–2663 [2020]

[28] Wu, Y.-H., Gao, S.-H., Mei, J., Xu, J., Fan, D.-P., Zhao, C.-W., Cheng, M.-M. - An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation - arXiv:2004.07054 [2020]

[29] Yan, Q., Wang, B., Gong, D., Luo, C., Zhao, W., Shen, J., Shi, Q., Jin, S., Zhang, L., You, Z. - Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review - Eur Radiol 30, 4381–4389 [2020]

[30] Ye, Z., Zhang, Y., Wang, Y., Huang, Z., Song, B. - COVID-19 Chest CT Image Segmentation – A multi-class COVID-19 segmentation network with pyramid attention and edge loss in CT images - IET Image Processing. 2021;1–10 [2021]

[31] Yu F., Zhu Y., Qin X., Xin Y., Yang D., Xu T. - A multi-class COVID-19 segmentation network with pyramid attention and edge loss in CT images - medRxiv 2020.03.12.20027185 [2020]

[32] Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W., Wang, X. - Deep Learning-based Detection for COVID-19 from Chest CT us-
ing Weak Label - medRxiv 2020.03.12.20027185 [2020]

[33] Zhou, T., Canu, S., Ruan, S. - An automatic COVID-19 CT segmentation network using spatial and channel attention mechanisms - arXiv:2004.06673 [cs, eess] [2020]
CoLe-CNN+: Context Learning - Convolutional Neural Network for COVID-Ground-Glass-Opacities Detection and Segmentation

Highlights

• Single and completely automated workflow for accurate lung segmentation, COVID lesions detection and segmentation.

• CNN producing four different segmentation mask that are combined in order to obtain a more accurate result.

• Specially developed post-processing tool to improve specificity, reaching 99.6%.

• Our method outperformed the state of the art reaching an average of 98.9% accuracy in COVID lesions segmentation.
Conflicts of Interest Statement

Manuscript title: CoLe-CNN+: Context Learning - Convolutional Neural Network for COVID-Ground-Glass-Opacities Detection and Segmentation

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers’ bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Author names:

The authors whose names are listed immediately below report the following details of affiliation or involvement in an organization or entity with a financial or non-financial interest in the subject matter or materials discussed in this manuscript. Please specify the nature of the conflict on a separate sheet of paper if the space below is inadequate.

Author names:

Giuseppe Pezzano - PhD student at Universitat de Barcelona, employed at Eurecat, salaried from Eurecat through the Vicente Lopez grant. His contract will be renewed on the basis of his results in research.

Oliver Diaz - Professor at Universitat de Barcelona.

Vicent Ribas - Group leader at Eurecat, internal supervisor of Giuseppe Pezzano.

Petia Radeva - Professor at Universitat de Barcelona, supervisor of Giuseppe Pezzano.
This statement is signed by all the authors to indicate agreement that the above information is true and correct (a photocopy of this form may be used if there are more than 10 authors):

| Author's name (typed) | Author's signature | Date       |
|-----------------------|--------------------|------------|
| Giuseppe Pezzano      | Giuseppe Pezzano   | 24/02/2021 |
| Vicent Ribas          | Vicent J. Ribas Ripoll | 25/02/21 |
| Petia Radeva          | Petia Radeva       | 24/02/2021 |
| Oliver Diaz           | Oliver             | 25/02/2021 |