The Improvement of Segmentation of Lung Pathologies and Pleural Effusion on CT-scans of Patients with Covid-19

D. Lashchenova^{*a*,*}, A. Gromov^{*b*,**}, A. Konushin^{*a*,***}, and A. Mesheryakova^{*b*,****}

^aLomonosov Moscow State University GSP-1, Moscow, 119991 Russia ^bThird Opinion Platform LLC, Skolkovo Innovation Center Territory, Moscow, 121205 Russia *e-mail: daria.laschenova@graphics.cs.msu.ru **e-mail: alexander.gromov@3opinion.ai ***e-mail: anton.konushin@graphics.cs.msu.ru ****e-mail: ceo@3opinion.ai Received October 10, 2020; revised October 20, 2020; accepted January 12, 2021

Abstract—In 2020 the outbreak of Covid-19 influenced lives of billions of people all around the globe and motivated governments of different countries to revisit the current situation with regards to public healthcare systems and to methods used in modern medicine. As the workload on radiologists and physicians increased, so did the demand on systems that automatically analyse medical images and detect pathologies. Many current computer vision papers assume that the solution would be integrated into a healthcare system. However improvement according to "classic" metrics like mAP or IoU does not necessarily mean improvement from the radiologist's point of view. In this paper we suggest that while calculating metrics, averaging should be performed not by all studies, but by different groups of studies, in order to be close to human perception of a quality of a segmentation. And that we should count the number of false positive components, found outside lungs, because the presence of such components is negatively perceived by radiologists. Also we propose a method that improves the segmentation of lung pathologies and pleural effusion according to the points given above.

DOI: 10.1134/S0361768821030063

1. INTRODUCTION

Covid-19 is a disease caused by the SARS-CoV-2 virus. Viral pneumonia is a common complication caused by Covid-19.

Early diagnosis of complications can reduce time and intensity of medical treatment. Usually the diagnosis is performed using computer tomography. CT is a type of radiography in which a three-dimensional image of the body structure is constructed from a series of cross-sectional images made along the axis. Despite the fact that CT-scanning is not the main instrument for Covid-19 identification, it is used for diagnosing complications, ranking patients by severity of complications, and monitoring the course of the disease. Radiologists detect various pathologies on CT-scans and evaluate their type and size. Then this information is used to determine a severity of a patient's condition and to select the appropriate treatment. There are five stages of pneumonia severity: CT-0, CT-1, CT-2, CT-3, CT-4. The stage is assigned, according to the percentage of lung opacity and is used for treatment selection and for determining whether the patient needs to be hospitalized. In severe cases pleural effusion, an excessive fluid between the layers of the pleura outside the lungs, may appear. It's important for physicians to detect this pathology and estimate its volume and fluid type.

The purpose of this study is to improve the system for automatic segmentation of lungs, pathologies of lungs, and pleural effusion on CT-scans used for estimating of the percentage of lung opacity and volume of pleural effusion.

2. RELATED WORK

As Covid-19 has been rapidly spreading, computer vision specialists have started to search for solutions to help radiologists.

Before the mass use of SARS-CoV-2 testing, some scientists were trying to find out the probability that a patient's pneumonia was caused by Covid-19, using solely CT scans. Thus, a lot of research on classification was done: Linda Wang [1] introduced COVID-Net, a neural network architecture that could classify Covid-19 in patients and distinguish between Covid-19-related and non-related pneumonia. Similarly, Lin Li proposed the neural network COVNet [2] for the same



Fig. 1. Examples of CT slices.

task. However such papers lost their relevance since Covid-19 testing became more accessible.

Currently, solving the task of pathology segmentation and pathology identification is required for creating a system for routing and monitoring a patients' condition. For example, Fei Shan proposed 3D-model VB-Net [3] for segmentation of pathologies, lung lobes and lung segments, training the model with human-in-theloop strategy. Parham Yazdekhasty and Ali Zindar [4] used a U-Net-like network with two decoders for prediction of pathology and lung classes and then combined results for improving detection of pathologies. In addition there are papers that cover both classification and segmentation problems such as performed by Amine Amyar [5].

3. DATASET

The dataset was provided by Third Opinion Platform [6]. It contained 938 studies of patients with Covid-19. Studies were stored in dicom format. Preprocessing was not performed. Examples of slices can be seen on Fig. 1.

10–20 chest slices from each study were annotated by radiologists. This gave us 18.673 images. The radiologists drew polygons surrounding lungs, pathologies of lungs, and pleural effusion.

The dataset was split into training and testing parts, containing 85% and 15% of studies respectively.

It is worth mentioning that radiologists tend to draw polygons with smooth boundaries, while networks give precise boundaries. In addition, some pathologies have uncertain boundaries, which makes creation of precise ground-truth masks harder. This results the fact that the quality metric values won't be close to theoretically ideal.

4. METRICS

4.1. General Metrics

Mean AP over test studies and IoU were used for model evaluation.

Model recall is calculated as the number of true positive pixels divided by the number of all positive pixels. Model precision is calculated as the number of true positive pixels divided by the number of pixels, predicted as positive. AP is calculated as the area under the precision-recall curve, where each point on the curve plot represents precision and recall of results with a threshold corresponding to the point.

IoU measures the overlap between ground truth and predicted areas:

$$IoU = \frac{TP}{FN + TP + FP}$$

It should be noted that mistakes in a small number of pixels lead to different impacts on these metrics in accordance with the number of true pixels in a ground-true mask. Moreover, studies that are missing certain classes are hard to evaluate using thosemetrics. It's impossible to calculate mAP on these studies; if a network doesn't find pixels of the certain class then IoU is either not calculable or we could put it to be 1, but if a network found at least one pixel of the class then IoU would dropto 0. And a small number of mistaken pixels would lead to a greater impact on the metrics in studies with a small number of true pixels.

This is why it was decided to calculate general metrics within 3 groups of studies: the mean number of false positive pixels is calculated in studies without pixels of a certain class, mAP and IoU are calculated separately on studies with number of pixelsless than and more than 10000.

4.2. Blob Metrics

During the dialog with radiologists, it was realized that sometimes even if a new model had an improvement according to general metrics, not only did they fail to notice the improvement, but claimed that the model worked worse. One of the reasons was thatthe model segmented blobs (connected components) of lungs and lung pathologies outside of real lungs, for example in the intestines. To detect this, connected components were computed for every class and for each component it was found whether it had intercep-



Fig. 2. Baseline architecture of the network.

tion with ground-truth lungs. If it doesn't, it means that it is a blob outside the lungs. As a metric, we use the number of blobs of different sizes outside the lungs: small (less than 100 pixels), medium (from 100 to 500 pixels), and large (more than 500 pixels). It's more important to decrease the number of large blobs, as they are more often noticed by radiologists.

5. PROPOSED METHOD

In this work we achieved the aim of reducing the number of blobs outside the lungs, by suggesting methods of decreasing the number of such components. The baseline is a U-Net-like [7] neural network. The architecture is shown on Fig. 2. It was trained on classes "background", "lung", "pathology", "pleural effusion" with cross-entropy loss function.

5.1. Prior Lung Segmentation

One of the approaches is to segment lungs and then segment pathologies inside masked lungs. Segmentation can also be performed on smaller images that will improve the quality of segmentation, because lungs occupy a larger part of a slice than pathologies. But in this case pleural effusion would be masked as it is not the part of lungs. Also this approach results in additional computational effort and time cost. One of the



Fig. 3. The example of a slice masked with filling constants corresponding to air (left) and tissue (right).

LASHCHENOVA et al.

Table 1. Comparison of the baseline method with methods with prior segmentation of "pathology" class. segm air – seg-mentation with filling constant corresponding to air, segm body – segmentation with filling constant corresponding to surrounding tissue

	FP (0)	IoU (pat, s)	IoU (pat, l)	mAP (pat, s)	mAP (pat, l)
baseline	575.6	0.413	0.603	0.656	0.856
segm air	641.5	0.402	0.596	0.626	0.840
segm body	783.6	0.374	0.584	0.604	0.833

Table 2. Model comparison by the number of blobs outside the lungs

	Lungs			Pathologies			
	small comp	med comp	big comp	small comp	med comp	big comp	
baseline	4718	207	78	1051	393	52	
rcca	5822	181	68	862	339	55	
aux fore	6085	120	21	1144	237	20	
aux same	6379	140	42	952	291	37	
rcca aux fore	6356	152	30	835	354	47	
rcca aux same	5788	128	31	736	330	41	

Table 3. Model comparison by the general metrics values for class "pathology"

	FP (pat, 0)	IoU (pat, s)	IoU (pat, l)	mAP (pat, s)	mAP (pat, l)
baseline	575.6	0.413	0.603	0.656	0.856
rcca	525.9	0.412	0.615	0.627	0.856
aux fore	635.0	0.408	0.610	0.636	0.850
aux same	547.1	0.413	0.604	0.643	0.848
rcca aux fore	618.8	0.400	0.620	0.638	0.856
rcca aux same	613.9	0.409	0.627	0.656	0.861

Table 4. Model comparison by the general metrics values for class "pleural effusion"

	FP (pl eff, 0)	IoU (pl eff, s)	IoU (pl eff, l)	mAP (pl eff, s)	mAP (pl eff, l)
baseline	961.6	0.383	0.632	0.590	0.803
rcca	407.9	0.320	0.636	0.556	0.807
aux fore	837.2	0.333	0.624	0.547	0.798
aux same	757.3	0.374	0.652	0.524	0.814
rcca aux fore	385.9	0.326	0.630	0.572	0.805
rcca aux same	698.7	0.369	0.660	0.596	0.812



Fig. 4. Architecture of the network with RCCA module.



Fig. 5. Architecture of the network with auxiliary loss.

questions of this approach is which constant to use for filling the masked area. On one hand it can be filled with value corresponding to air, but on other hand with value, corresponding to surrounding tissue. Both ways were tried. Examples of different constant usage can be seen on Fig. 3. tion within the whole image, so the network could better differentiate pathologies and lungs from similar structures, using the context from the whole image. In this work the rcca module is placed between the encoder and the decoder of the baseline network as shown in Fig. 4.

5.2. Inserion OF RCCA Module

This module was introduced in the paper CCNet: Criss-Cross Attention for Semantic Segmentation [8]. The main idea of the module is to distribute informa5.3. Adding of Auxiliary Loss

Also it was suggested that inserting an additional head after the encoder or the RCCA module and computing coarse segmentation on a smaller size would



Fig. 6. Architecture of the network with RCCA module and auxiliary loss.

PROGRAMMING AND COMPUTER SOFTWARE Vol. 47 No. 4 2021



Fig. 7. Example of work on the slice with pleural effusion.

prevent neural networks from finding big components in wrong places. Two types of auxiliary losses were compared: the same loss that was used in the decoder, but on the label map of a smaller size, and binary cross entropy loss that is used to distinguish the foreground from the background.

6. RESULTS

In the tables below (classname, 0) refers to studies without specific class, (classname, s) to studies which have less than 10000 pixels in ground truth mask, (classname, l) to studies which have less than 10000 pixels in ground truth mask. Baseline refers to baseline method, rcca in name refers to methods that use rcca, aux means presence of auxiliary loss: aux fore for binary cross entropy after sigmoid activation that distinguishes the foreground from the background, aux same for cross-entropy after softmax function for prediction of coarse masks.

6.1. Experiments with Prior Lung Segmentation

Experiments were made with models that predicted class "pathology" on images with masked lungs.

As seen from Tables 1-4, prior segmentation gives worse results in both studies with pathologies and studies without pathologies, because it predicts more false positive pixels.

It's noteworthy that the choice of a filling constant can influence the result. It would be a good idea to pay attention to it.

6.2. Experiments with RCCA Module and Auxiliary Loss

Experiments were made with models that predicted classes "background", "lung", "pathology", and "pleural effusion".

First of all, the influence of these approaches on blobs detected outside lungs is analysed. Results are provided in the Tables 1-4.

Then the analysis of the influence on general metrics of segmentation quality for the classes "pathology" (Tables 1-4) and "pleural effusion" (Tables 1-4) is performed.

Comparing the values of metrics, we conclude that the usage of auxiliary loss reduces the number of blobs outside the lungs and slightly decreases the quality in studies with pathologies of small size, but increases the quality in studies with pathologies of big size. Also, despite the decrease of the segmentation quality due to the usage of the RCCA module, if the auxiliary loss is applied afterwards, it gives benefits to the studies with big size pathologies and reduces the number of blobs outside, while slightly decreasing values of general metrics in studies with a small number of pathologies. Examples of how the model works can be seen on Fig. 7. In the upper row in the left corner radiological assessment is shown, on the right corner there is the model prediction. In the bottom row there is an original slice with different display settings.

CONCLUSION

To evaluate automatic systems used by people, it is necessary to apply not only "classic" metrics of segmentation, but also metrics that may detect humanrelevant errors. Usually when comparing different models, there is no single model that achieves better metrics over all other models. Therefore it is necessary to seek a compromise.

REFERENCES

- Wang, L., Lin, Z.Q., and Wong, A., Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, *Sci. Rep.*, 2020, vol. 10, no. 1, pp. 1–12.
- 2. Li, L., et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, *Radiology*, 2020, art. 200905.
- 3. Shan, E., et al., Lung infection quantification of covid-19 in ct images with deep learning, 2020. arXiv:2003.04655.
- 4. Yazdekhasty, P., et al., Bifurcated autoencoder for segmentation of COVID-19 infected regions in CT images, 2020. arXiv:2011.00631.
- Amyar, A., et al., Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation, *Comput. Biol. Med.*, 2020, vol. **126**, p. 104037.
- 6. "Third Opinion Platform" Limited Liability Company. https://thirdopinion.ai/.
- 7. Ronneberger, O., Fischer, P., and Brox, T., U-net: convolutional networks for biomedical image segmentation, in *Proc. Int. Conf. on Medical Image Computing and Computer Assisted Intervention*, Springer, 2015, pp. 234–241.
- 8. Huang, Z., et al., Ccnet: criss-cross attention for semantic segmentation, *Proc. IEEE Int. Conf. on Computer Vision*, Seoul, 2019, pp. 603–612.