Image Segmentation as an Instrument for Setting Attention Regions in Convolutional Neural Networks for Bias Detection Purposes

Bojana Velichkovska, Danijela Efnusheva, Marija Kalendar and Goran Jakimovski
Faculty of Electrical Engineering and Infomation Technologies, "SS. Cyril and Methodius University" in Skopje,
Rugjer Boshkovikj Str. 18, Skopje, N. Macedonia
{bojanav, danijela, marijaka, goranj}@feit.ukim.edu.mk

Keywords: Artificial Intelligence, Deep Learning, Medical Image Processing, Convolutional Neural Networks, Attention Regions, Lung Segmentation.

Abstract: Convolutional neural networks (CNNs) are constantly being used for medical image processing with increased application in publicly available datasets and are later being actively applied in medical practice. Therefore, since patient lives are at stake, it is important that the functionality of the neural network is beyond reproach. In this paper, due to dataset availability, we present two lung segmentation approaches using traditional image processing and deep learning methodologies; these approaches can later be used to focus a CNN for image segmentation and classification tasks, with implementations spanning everything from disease diagnosis to demographic and bias analysis. The aim of this paper is to provide a framework for segmentation in medical images of the chest cavity, as a way of applying attention regions and localizing sources of bias in images. Both of the proposed segmentation tools, the traditional image approach using computer tomography scans and the CNN applied to chest X-rays, provide excellent lung segmentation comparable to popular methods in the image processing sphere. This allows for an all-encompassing application of the developed methodology regardless of different image formats, therefore making it widely applicable in setting attention regions for CNNs.

1 INTRODUCTION

Deep learning approaches based on convolutional neural networks (CNNs) have allowed computers to achieve excellent results in the field of computer vision. Namely, CNNs have found application in tasks like image classification, object detection, and semantic classification [1][2]. Furthermore, CNNs have significantly contributed in medical image processing [3][4][5]. Researchers have successfully applied CNNs in many medical applications, such as tumor classification [6], detection of skin lesions [7], heart anomalies [8][9], etc.

It is often necessary to focus the attention of the network, namely to restrict the recognition of the network to a specific region in the image. This region is known as a region of interest (ROI) [1] or attention mask, and it is given as an input to a CNN in order to provide the focus. However, before being applied for that purpose, the ROI must be detected and properly defined. Naturally, it is easier to use already existing methodologies and defined approaches as means to extrapolate the ROI in an image.

One of the standard approaches of introducing a ROI as an input for CNNs is by assuming a fixed, rectangular ROI alike a bounding box. This ROI can then be cropped and used as a separate input to the CNN. However, there are several limitations to this approach. One important constraint to consider is that the ROI assumes a rectangular shape, and as such it is not applicable to problems where the investigated elements have an arbitrary shape, which is often the case when working with medical images. An additional issue is that with this approach certain background information, that might be essential for understanding the context of the features obtained, will be ignored due to the selective cropping.

On the other hand, irregularly-shaped ROI offer the ability to select all sections of interest, background included, and feed them to the network for the analysis. For example, with a specific image in mind, for the purposes of one study the required accent can be on the bones, whereas another might require the lungs.

Therefore, in this paper, we investigate two methodologies – traditional image processing and deep learning – in order to perform lung segmentation on two different types of medical images: chest X-rays and computer tomography (CT). The purpose is to analyze the quality of the obtained segmentation results and understand whether they could be used as attention regions in CNNs in bias analysis of medical images. The topic of bias has been widely investigated in the past few years [10]. Reported cases of bias include medical personnel, medical datasets, and medical AI-based applications [11]. With awareness levels rising, researchers have begun to analyze the presence of bias in medical images [12], with nearly 100% accuracy in gender bias and 90% accuracy in racial bias. Curiosity arises in the case of racial bias and the sources which are indicative of its presence from mere chest X-rays, therefore we wish to develop an instrument which would allow investigation of the elements which the network uses to detect it.

Due to the type of data available, we center our approach around evaluation of lung segmentation as a tool. However, the developed pipeline is applicable to other forms of segmentation (e.g., bone, soft tissue), should the data required become available. In order to evaluate the quality of the results, we compare our proposed CNN to the U-Net network [13] which is one of the standards in the field of image processing.

This paper is organized as follows. Section two provides the dataset, technology, and evaluation metrics used. Section three gives an overview of the results and in section four we conclude the paper.

2 MATERIALS AND METHODS

The pipeline for lung segmentation used in this research consists of three stages: data preprocessing, model training, and model evaluation. For that purpose, in this section we describe the datasets used in the paper, then address the technology used, and finally we provide the evaluation metrics used for evaluating the performance of the network from the results obtained.

2.1 Datasets

For the purposes of this research, we use two different image datasets. The first dataset contains chest X-rays collected by radiologists at two clinics (Montgomery and Shenzhen) [14]. The data encompasses chest X-rays from approximately 600 patients with tuberculosis. Additionally, following anatomical landmarks, the X-rays were also accompanied by corresponding lung segmentations. The second dataset contains CT images of patients suffering from pulmonary fibrosis created by the Radiological Society of North America [15].

The reason for the different data sets lies in the format. Namely the research focuses on X-ray images and CT images, as to offer a comprehensive approach for lung segmentation methods and applications on different medical imaging formats.

Both datasets are represented in the DICOM format. DICOM is a medical image processing information communication and management standard which is used to store, exchange, and transmit medical images. The standard includes protocols for exchange, compression, and 3D visualization of results for multiple medical procedures, like magnetic resonance, radiography, computed tomography, etc.

When it comes to the first dataset, one record in this standard consists of a set of pixels that represent a static position of the chest. On the other hand, when it comes to the second dataset, the representation of CT images is more complex, in that, it includes a set of static positions taken at different depths (or sections) of the body in order to create an overall image of the chest cavity. This format is widely used due to its advantageousness; namely, the DICOM standard offers high quality of the stored information, i.e., the images have higher dimensionality compared to what the human eye can perceive.

2.2 Traditional Image Processing Methodology

For the traditional image processing, we relied on the Hounsfield scale, which is a quantitative scale for describing radiodensity. The scale provides a clear overview of the information embedded in the image, or simply put, by applying a simple range of values one can understand the tissue displayed in the image. A detailed overview of the Hounsfield units is given in Table 1, where, e.g., a unit measure above 1000 means the image contains bone, calcium, or metal, whereas a unit measurement below -1000 signifies air, and so on.

However, the results obtained after applying the Hounsfield units scale can generate masks which contain artefacts (anomalies in the image, e.g., missing pixel values in the center of a mass whether it be bone or an organ). In such cases, we apply a single round of basic morphological operations, in our case, erosion and dilation, to compensate for the errors. Dilation adds pixels to the boundaries of objects in an image, which fills out the potential anomalies in the center of the lungs. Erosion removes pixels on object boundaries, meaning it decreases the volume expansion of the lungs (originating from the dilation).

Table 1: Hounsfield units.

Unit Measure	Representation of
> 1000	Bone, calcium, metal
100 to 600	Iodinated CT contrast
30 to 500	Punctate calcifications
60 to 100	Intracranial hemorrhage
35	Gray matter
25	White matter
20 to 40	Muscle, soft tissue
0	Water
-30 to -70	Fat
< -1000	Air

2.3 CNNs

CNNs as a segmentation tool, are a structure of one or more convolutional layers, often followed by sampling layers and one or more deconvolutional layers. The input and the output layers of a CNN serve the purpose of defining the functionality of the CNN. Namely, the input layer defines what the network requests as input data, and the output layer defines what the network will provide as a result of the input. On the other hand, the type of hidden layers structured in the CNN and the way these layers are connected define the behavior of the network, or rather what the network will observe, compute and learn from.

In order to segment the lung from the chest X-rays we use a CNN consisting of four convolutional and four deconvolutional layers. The last three of the convolutional layers are followed by an undersample layer, and the first three deconvolutional layers are preceded by oversampling layers. The final layer is followed by a softmax activation function that provides the segmentation result.

2.4 Evaluation Metrics

The choice of metrics when evaluating the method’s performance is important when training deep learning models [16]. The reason for this is because the same model can give different results if the analysis is performed using different metrics; namely, one evaluation metric can suggest good results, whereas the model is actually underperforming when another evaluation metrics is considered.

One of the most often used evaluation metrics is accuracy. When it comes to segmentation, accuracy as a metrics is indicative of pixel-wise classification. Now, the negative aspect of this metrics comes to light when working with significantly imbalanced data, which in the case of segmentation problems can often be the case. In cases as this, the value of accuracy can easily reach the actual percentage representation of the class that dominates the data set, while the model itself is still a weak classifier because it knows how to recognize only the dominant class.

As a result, a better metrics for evaluating segmentation models is the confusion matrix and the all-encompassing metrics which can be extrapolated from it. In this paper, we use the dice coefficient, given in (1). The values used to calculate the dice coefficient are as follows: true positives (TP), false positives (FP), and false negatives (FN). A TP is an outcome where the model correctly predicts the positive class. A FP is an outcome where an instance is predicted positive when it is actually negative, whereas a FN is an outcome when an instance is predicted negative when it is actually positive.

$$\mathrm{diceScore} = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

The dice coefficient calculates the similarity of two samples, and its values can range from 0 to 1, with 0 indicating no overlap between two segmented areas, whereas 1 represents full overlap between the proposed and the true segmentation areas. Therefore, the higher the value of the dice coefficient the better the segmentation.

3 RESULTS

The results obtained are divided into two separate groups: results from the traditional image processing obtained from the CT scans and results from CNNs obtained from the chest X-rays.

3.1 Traditional Image Processing Methodology

The results obtained through the traditional image processing in the CT scans can be observed in Figure 1. In the observed masks artefacts can be seen in several images. This means that single dilation cannot fix the artefact, therefore in the

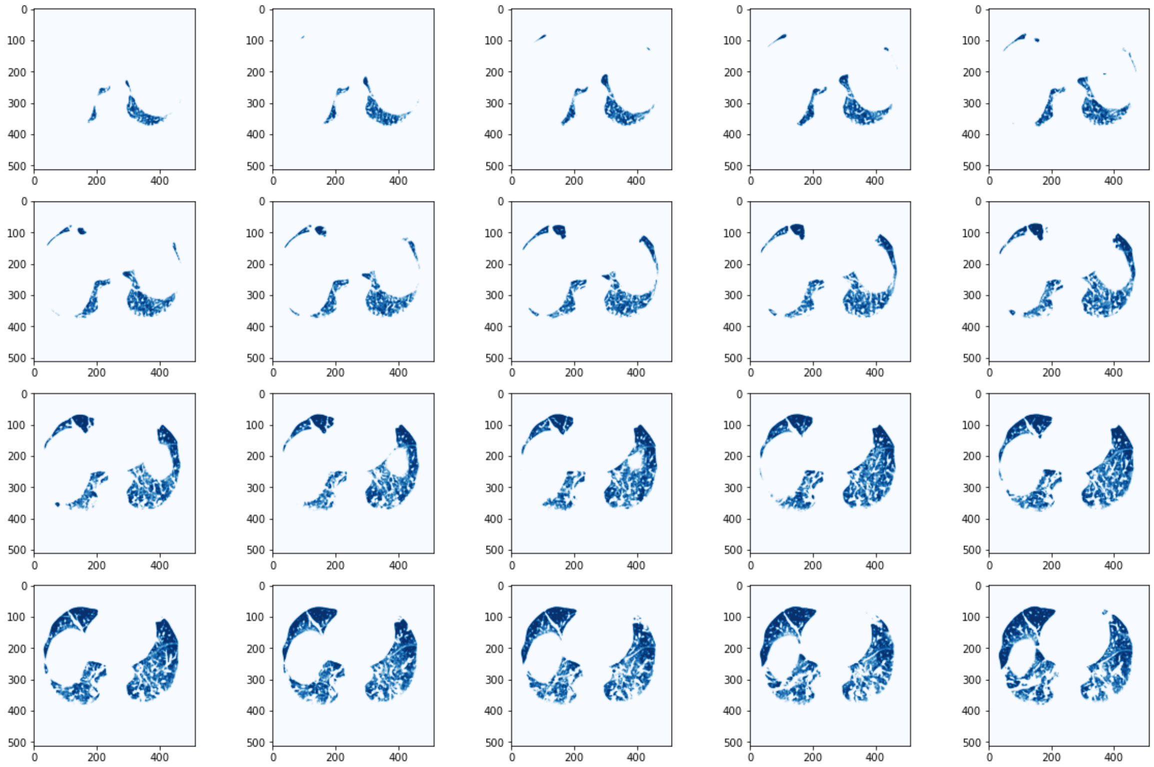

Figure 1: Results obtained from traditional image processing using the Hounsfield scale from all slices of a randomly selected CT scan.

future we can expand these results and test the number of dilation-erosion repetitions required to fix the artefacts from the segmentation errors based on the Hounsfield scale.

The result from the final segmentation can be seen in Figure 2. This is a three-dimensional rendering obtained by merging all the sections (layers) from the CT scans into one. The rendered result shows that using the Hounsfield scale provides excellent lung segmentation results, in spite of the small artefacts and excess selected regions, which can be seen in the upper left and bottom right corners of the rendition.

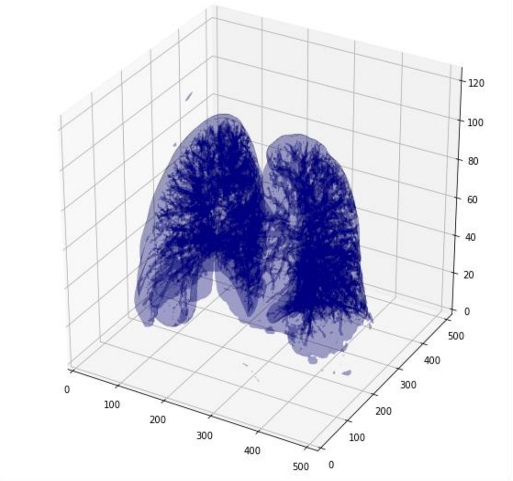

Figure 2: Three-dimensional rendering of the segmented lungs obtained from all slices in a single CT scan.

3.2 CNNs

We investigated two different CNNs. One is our proposed network consisting of four convolutional layers and four deconvolutional layer. The deconvolutional layers are concatenated with the convolutions for the purpose of eliminating noise, whereas the other is U-Net. Both models were trained and evaluated on the chest X-ray dataset.

The difference between the U-Net network and our network is the number of layers which encode the images. Namely, our proposed method uses less layers to analyze the input, whilst it provides comparable results to U-Net. We are using a two-layered concatenation, whereas U-Net concatenates three input layers. This makes the proposed model faster to train compared with U-Net, but the overall performance is only marginally affected.

Both of the networks were trained on the same portion of the Montgomery and Shenzhen chest X-ray dataset, and both were optimized using Adam [17]. During the training stages, the monitored metrics was the dice coefficient, which was later also used to evaluate the overall performance. The results for the U-Net are given in Figure 4, while the results for our proposed method are given in Figure 3. In both figures, the first column of images is the predicted segmentation, the second column are the actual labels, whereas the third column describes the difference between the two. The difference map contains four separate colors, each of which depicts a certain aspect of the confusion matrix. The light pink indicates lung segments which were correctly identified, whereas the black shows the correctly

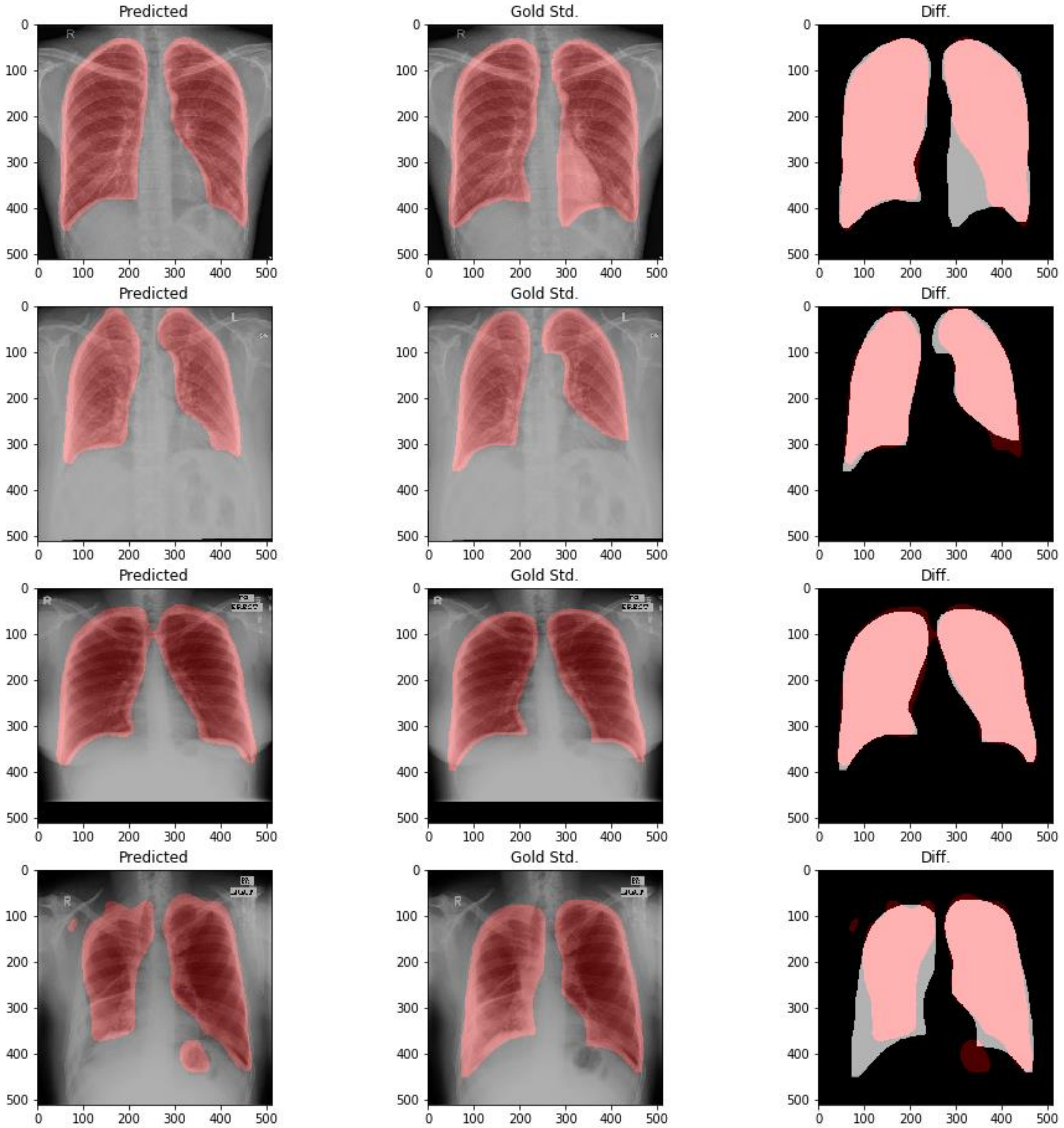

Figure 3: Results obtained from the proposed CNN. Each row represents a different chest X-ray. For the columns: the first column displays the chest X-ray overlapped with the segmentation mask predicted by the CNN, the second column displays the chest X-ray overlapped with the segmentation label (or rather the annotation provided by the dataset creators), and the third column shows the difference between the predicted and the actual masks.

removed background elements. The light gray shows the sections which should have been identified as lungs but were instead classified as background, and the burgundy shows where the model misclassified the background as lung tissue.

Major differences in the results can be noted in the last row of X-ray images in the result figures for both datasets. We can see that the proposed CNN struggles with properly segmenting a section of the left lung, which is not the case with the results obtained from U-Net. Therefore, it can be concluded that one of the limitations of the proposed CNN is the ability to differentiate the lung in instances where clouding of the lower lung is present. However, if we observe the first row of both figures, it can be noted that there are cases where the U-Net also has a difficult time identifying lower sections of the lung.

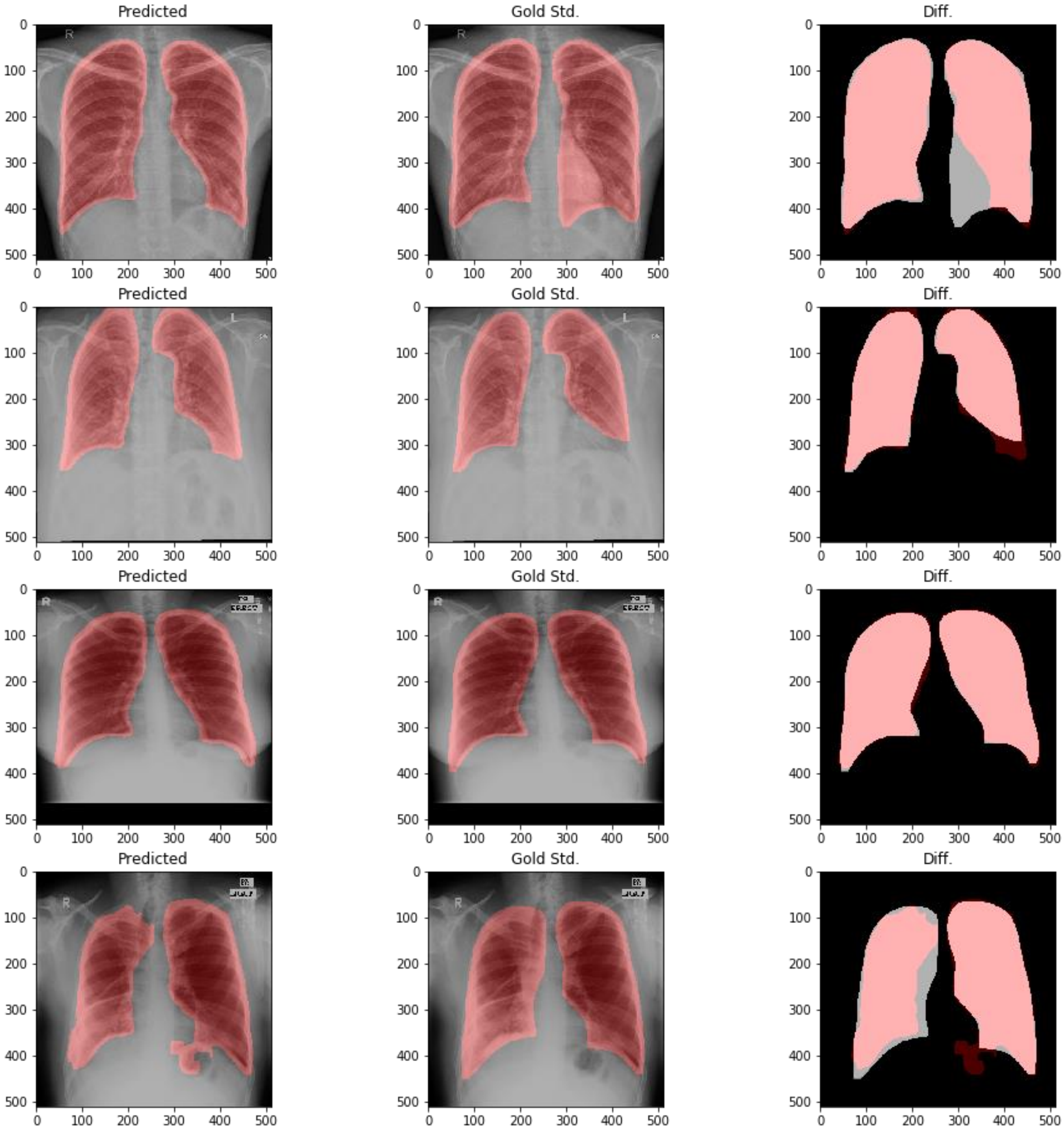

Figure 4: Results obtained from U-Net. Each row represents a different chest X-ray. For the columns: the first column displays the chest X-ray overlapped with the segmentation mask predicted by the CNN, the second column displays the chest X-ray overlapped with the segmentation label (or rather the annotation provided by the dataset creators), and the third column shows the difference between the predicted and the actual masks.

The dice coefficient for the proposed CNN is 0.95, while for U-Net it is 0.97, making the results between the two CNNs comparable, as can be seen in the values of the dice coefficient and the selected samples, with the added advantage of faster training time in the case of our proposed approach.

4 CONCLUSIONS

With this paper we proposed two separate methods for lung segmentation obtained from two different medical imaging technologies represented by a single format. Both of the proposed segmentation tools provide excellent lung segmentation in their corresponding datasets, which provides two different approaches in obtaining proposed attention regions for CNNs in classification and segmentation problems requiring focus on image sections. Since the results are comparable to a widely-used and renowned U-Net, the proposed instrument is likely to be an effective tool for focusing CNNs in future research. Namely, the overall idea is to utilize the obtained results for investigating presence of bias in imaging datasets, as well as understanding where exactly that bias originates from. Therefore, as a future step, this research will be expanded into segmenting different aspects of the chest cavity (e.g., bones, soft tissue, etc.) and applying focus on those areas in order to determine which elements of the image (and therefore which parts of the human body) contribute to presence of bias the most.

REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection Region Proposal Networks,” Advances in Neural Information Processing Systems, vol. 28, 2015.

[2] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation.,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[3] P. Dutta, P. Upadhyay, M. De, and R. G. Khalkar, “Medical Image Analysis using Deep Convolutional Neural Networks: CNN Architectures and Transfer Learning,” 2020 IC on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020.

[4] P. Bir, and V. E. Balas, “A Review on Medical Image Analysis with Convolutional Neural Networks,” 2020 IEEE IC on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2020.

[5] P. Kalyani, S. Srivastava, A. Reddyprasad, R. Krishnamoorthy, S. Arun, and S. Padmapriya, “Medical Image Processing from Large Datasets Using Deep Learning,” 2021 3rd IC on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021.

[6] J. E. A. Ovalle, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. Á. Guevara-López, “Representation learning for mammography mass lesion classification with convolutional neural networks”, Computer Methods and Programs in Biomedicine, vol. 127, pp. 248-257, 2016.

[7] B. Shetty, R. Fernandes, A.P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. Lakshmanna, “Skin lesion classification of dermoscopic images using machine learning and convolutional neural network,” Sci Rep, vol. 12, no. 1, pp. 18134, 2022.

[8] S. P. Jillella, C. Rohith, S. Shameem and P. S. S. Babu, “ECG Classification For Arrhythmias using CNN & Heart Disease Prediction using Web application,” 2022 First IC on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022.

[9] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, et al., “Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation,” IEEE Transactions on Medical Imaging, vol. 37, no. 2, pp. 384-395, 2018.

[10] S. Romano, D. Fucci, G. Scanniello, M. Teresa Baldassarre, B. Turhan, and N. Juristo, “Researcher Bias in Software Engineering Experiments: a Qualitative Investigation,” 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Portoroz, Slovenia, 2020.

[11] J. A. Sabin, “Tackling Implicit Bias in Health Care”, New England Journal of Medicine, vol. 387, no. 2, pp. 105-107, 2022.

[12] J. W. Gichoya, I. Banerjee, A. R. Bhimireddy, and et al., “AI recognition of patient race in medical imaging: a modelling study,” in The Lancet Digital Health, vol. 4, no. 6, pp. 406-414, 2022.

[13] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” Medical Image Computing and Computer-Assisted Intervention, Springer, vol. 9351, pp. 234-241, 2015.

[14] S. Jaeger, S. Candemir, S. Antani, Y. X. Wáng, P. X. Lu, and G. Thoma, “Two public chest X-ray datasets for computer-aided screening of pulmonary diseases,” Quant Imaging Med Surg, vol. 4, no. 6, pp. 475-477, 2014.

[15] E. Colak, F. C. Kitamura, S. B. Hobbs, and et al., “The RSNA Pulmonary Embolism CT Dataset,” Radiology: Artificial Intelligence, vol. 3, no. 2, p. e200254, 2021.

[16] M. M. Jawaid, R. Rajani, P. Liatsis, C. C. Reyes-Aldasoro, and G. Slabaugh, “Improved CTA Coronary Segmentation with a Volume-Specific Intensity Threshold,” Medical Image Understanding and Analysis, pp. 207-218, 2017.

[17] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, 2014.