

Development of a lung segmentation algorithm for analog imaged chest X-Ray: preliminary results

Matheus A. Renzo
Data Scientist
Murabei Data Science
São Paulo, Brasil
m.renzo@murabei.com

Natália Fernandez
Research Associate
Federal University of Rio de Janeiro
Rio de Janeiro, Brasil
nataliaffernandez@gmail.com

André A. Baceti
CTO
Murabei Data Science
São Paulo, Brasil
a.baceti@murabei.com

Natanael Nunes de Moura Junior
Adjunct Professor
Federal University of Rio de Janeiro
Rio de Janeiro, Brasil
natmourajr@lps.ufrj.br

André Anjos
Scientific Researcher
Idiap Research Institute
Martigny, Switzerland
andre.anjos@idiap.ch

Abstract—Analog X-Ray radiography is still used in many underdeveloped regions around the world. To allow these populations to benefit from advances in automatic computer-aided detection (CAD) systems, X-Ray films must be digitized. Unfortunately, this procedure may introduce imaging artefacts, which may severely impair the performance of such systems.

This work investigates the impact digitized images may cause to deep neural networks trained for lung (semantic) segmentation on digital x-ray samples. While three public datasets for lung segmentation evaluation exist for digital samples, none are available for digitized data. To this end, a U-Net-style architecture was trained on publicly available data, and used to predict lung segmentation on a newly annotated set of digitized images.

Using typical performance metrics such as the area under the precision-recall curve (AUPRC), our results show that the model is capable to identify lung regions at digital X-Rays with a high intra-dataset (AUPRC: 0.99), and cross-dataset (AUPRC: 0.99) efficiency on unseen test data. When challenged against digitized data, the performance is substantially degraded (AUPRC: 0.90).

Our analysis also suggests that typical performance markers, maximum F1 score and AUPRC, seems not to be informative to characterize segmentation problems in test images. For this goal pixels does not have independence due to natural connectivity of lungs in images, this implies that a lung pixel tends to be surrounded by other lung pixels.

This work is reproducible. Source code, evaluation protocols and baseline results are available at: <https://pypi.org/project/bob.ip.binseg/>.

Index Terms—Neural Network, U-Net, Lung Segmentation, Analog x-ray

I. INTRODUCTION

Computer-aided detection (CAD) algorithms were developed in the past years for tissue segmentation and disease diagnosis [1]–[4], with the objective of helping healthcare workers in screening and treatment of diseases. Since the beginning of the COVID-19 pandemic, the demand for this type of solution has increased further, as in the early days X-

ray and tomographic findings were used as a main diagnostic tool [5]–[7].

Despite recent interest due to the COVID-19 pandemic, Chest X-Ray (CXR) is also an important technique for screening Tuberculosis (TB) patients [8]. Clinical symptoms for TB are non-specific and, even in places where it is highly endemic, most of people experiencing those symptoms do not actually have TB [8]. TB, although being completely treatable in most cases, claims almost 4000 lives every year, being inversely correlated with Gross Domestic Product (GDP), and having a direct correlation with under nutrition [9]. This basically means that underdevelopment/developing countries are particularly susceptible to TB mortality.

Different methods were described for lung segmentation in digital CXR, but only a handful of them were challenged against images acquired from analog films [10]. This is particularly important in under developed regions where CXR equipment is old, still relying on analogical processes. In those cases, to allow CAD usage, one must first digitize the X-Ray films. Digitization can be accomplished using different devices such as photo cameras, specialized scanners, or even mobile devices. This digitization process is typically referred as “imaging”.

The use of digital X-Ray may introduce artefacts [11], and the same is true for imaging an analog X-Ray films. The distortions on the lung’s image could occur due to device resolution, image corrections, image compression, acquisition setup, and user capture procedures [12]. When the imaging process is done by a camera or mobile device, it is also possible that an important part of the image does not correspond to the X-Ray *per se*. Background such as walls, furniture, uncovered negatoscope bits, and other unpredictable elements are typically present. Identifying correctly the lung area can reduce the noise transferred to the CAD by these artefacts, which may increase effectiveness of CAD systems [2], [13].

This work shows preliminary results of a lung segmentation

algorithm based on a U-Net architecture [14]. It is part of a broader project which aims to develop a CAD to detect Tuberculosis (TB) in both digital and digitized images. Section II discusses related work on lung image segmentation of digital x-rays. Section III explains the proposed approach used in this study. Section IV describes the methodology used in the training and evaluation process. In Section V, the experimental results obtained using the proposed method are analyzed. Lastly, the results are discussed and a conclusion is drawn in Section VI.

II. RELATED WORK

In the last few decades, many research groups have worked on the task of analyzing CXR images, resulting in various methods proposed for lung-segmentation [15]. For example, in [16] Canny edge detection was used for detecting the lung boundaries. However, in the past decade, deep neural networks seem to have been given more attention as the preferred method when doing lung segmentation of digital x-rays: In [17], two deep convolutional neural network models were used, separating the tasks in four steps main steps: (1) image acquisition, (2) initial segmentation, (3) reconstruction and (4) final segmentation. Numerous other works tend to use a type of U-Net architecture based neural network, as in [18] with a variational encoder and decoder, or with attention modules such as in [19] or [20].

III. PROPOSED APPROACH

A. Network Architecture

To produce accurate lung segmentation masks on digital chest X-rays, pixel-wise dense predictions, regarding whether the pixels is part of the lung or not, are necessary. To achieve this, Fully Convolutional Networks (FCNs) are widely used and known to outperform other methods at this specific task [20] and the U-Net could be considered one of the most utilized FCNs architecture for segmentation in the field of medical imaging [21].

Having said that, a MobileV2Net U-Net model [22] was used for this exploratory work. This network incorporates the effectiveness of a U-Net model [14] along with the computational speed of a MobileV2Net [23], using a VGG-16 [24] backbone. The U-Net gets its name from its symmetric U-shape, having two paths, an encoder-contraction path, and a decoder-expanding path. The contraction path is where the context of the image is obtained through various convolutional and max-pooling operations (to down sample the size of feature map and contain less parameters in the network). In the symmetric expanding path on the right side, the image is up sampled using transposed convolution operations, in order to get pixel-wise predictions with the same size as the input image.

The MobileNetV2 is built on the idea that convolutional layers, though essential to computer vision tasks yet computationally costly, can be replaced by bottleneck residual blocks. The work of a convolutional layer is then split into 3 parts. First an 1x1 convolution that expands the number of channels

(thereby called an expansion layer), which is followed by a depth-wise convolution layer that filters the inputs, and lastly followed by a 1x1 projection layer that reduces the number of channels.

B. Binary Cross Entropy and a Soft Jaccard Loss

The loss selected is a combination of a binary cross entropy and soft Jaccard loss implemented in [25]. Binary cross entropy loss, denoted as H , can be written as (Eq. 1):

$$H = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \quad (1)$$

Where y is the label (1 for where the pixel is classified as lung and 0 otherwise) and $p(y)$ is the predicted probability of the pixel being classified lung, for N number of pixels. The Jaccard index is a similarity measure between a finite number of sets. For two sets A and B , it can be defined as (Eq. 2):

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2)$$

For pixels in an image, the expression can be adapted to (Eq. 3):

$$J = \frac{1}{N} \sum_{i=1}^N \frac{y_i \cdot p(y_i)}{y_i + p(y_i) - (y_i \cdot p(y_i))} \quad (3)$$

Ultimately, the combination of both losses, where in our experiments an alpha of 0.7 was used, is obtained as following (Eq. 4):

$$L = \alpha H - (1 - \alpha)(1 - J) \quad (4)$$

C. Adabound Optimizer

The optimizer selected is a variant of the Adam and AMS-Grad, named AdaBound [26], which uses dynamic bounds on learning rates for smoother transitions. The lower and upper bound are initialized as zero and infinity respectively, eventually converging to a constant final step size. It is designed to be more resilient to extreme learning rates. AdaBound is an optimizer that behaves like Adam at the beginning of training - achieving a rapid training process, and gradually transforms to SGD at the end, which tends to generalize better compared to adaptive methods.

D. F1 Score and AUPRC Performance Metrics

F1-Score (also known as Dice score) is a common used metric to evaluate the performance of a binary segmentation models[22]. It is computed as the harmonic mean of the precision and recall (Eq. 5), ranging from 0 to 1 (where 0 is the worst possible score and 1 being the best score):

$$F1 = 2 \cdot \frac{[precision] \cdot [recall]}{[precision] + [recall]} \quad (5)$$

The second metric used was the AUPRC. It comes from a Precision-Recall curve, which is a plot of the Precision

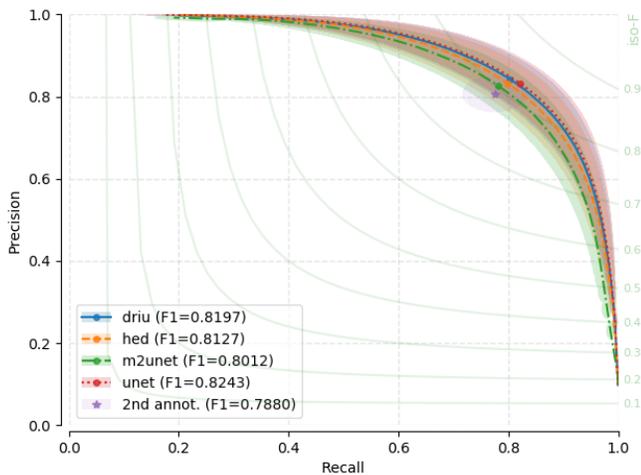


Fig. 1. Example of Precision-Recall curves on the Binary Segmentation Benchmark Package for Bob. Extracted from Laibacher and Anjos [27].

(on the y-axis) and the Recall (on the x-axis) for distinct probability thresholds (Fig. 1). A perfect model would be depicted as a single point at the coordinate of (1,1). A high-performing model is illustrated by a curve that bows towards that point of coordinate (1,1). It's a very effective plot to identify imbalanced models. The area under the curve (AUC), summarizes the PR curve with a range of probability thresholds into one single value (AUPRC). A perfect model would have a AUPRC score of 1.0.

IV. METHODS

A. Datasets

Erect anteroposterior chest (AP) x-ray images were collected from three public datasets: The Japanese Society of Radiological Technology (JSRT) dataset [28], the Montgomery County [29] and Shenzhen Hospital dataset [29] (Tab. I, Fig. 2). The data for all three datasets were split randomly, with 70% of the data reserved for training, 10% for validation, and 20% for testing. This specific data split was done in order to compare our proposed approach to other high-performing models ([30], [20], and [19]) that split these same datasets in similar fashion.

TABLE I
SUMMARY OF THE DATASETS WITH NUMBER OF IMAGES, MEDICAL DIAGNOSIS AND AREA ANNOTATION.

Dataset	N Images	Diagnosis	Annotation
JSRT	154	Lung nodules	Heart, clavicles and lungs
Montgomery	80	Tuberculosis	Lung
Shenzhen	672	Tuberculosis	Lung ^a

^a 566 images of the dataset were annotated [31].

B. Training

The training process was done on a GPU, using the Binary Segmentation Benchmark for Bob [27] package written in Python. Since the number of images in each dataset was



Fig. 2. Examples from all three datasets. Left: Montgomery County; middle: Shenzhen; right: JSRT.

limited, the model was initialized using pre-trained weights from ImageNet [32]. For the purpose of setting these initial experiments and their results as a baseline, the networks were trained for a fixed number of 1000 epochs, with a learning rate of 0.001, a final learning rate of 0.1, and a batch size containing 8 images, following the pattern of the other models trained on the aforementioned package. Images were resized to 512 by 512 pixels before training (following the same process found in [30] and [20]).

In order to explore the generalization potential of the previously described network model, all three datasets were used independently of each other. Subsequently, tests were conducted across all datasets and models, in addition to tests performed on the analog x-ray photo dataset consisting of 27 images.

C. Imaging procedure and App

Images were acquired using a mobile application (Imaging Manager) developed for this purpose by Murabei Data Science. The app does not allow the use of flash, and applies a timer of 3 seconds to enhance photo stability. It also has a mask that facilitates the centering of the x-ray image through red lines (Fig. 3). Users are instructed to align the vertical line to the spine and the horizontal one with the clavicles.

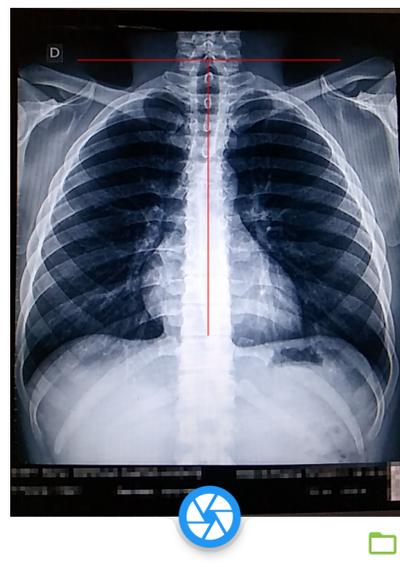


Fig. 3. Screenshot of the app Imaging Manager. Users were instructed to align the spine in the vertical line and the clavicle on the horizontal one.

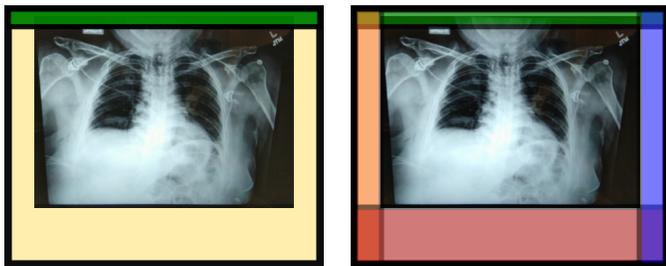


Fig. 4. Schematics of the x-ray photo acquisition. Users were instructed to cover the exposed remaining light box (left) with a opaque textile (right). Colors represent different textile rags.

The x-rays films were placed in an x-ray negatoscope and the remaining surface was covered by an opaque textile to block the light (Fig. 4).

Users were also instructed to turn-off ambient light and block other external sources such as windows and open doors. The cellphone was placed on a tripod during the image acquisition to reduce shaking effects.

D. Annotation of the lung segmentation on photos

The annotations for the lung segmentation task of 27 photos were performed by a physician using MakeSense.ai¹, a free tool for this type of workflows. He was asked to identify the lung area by drawing polygons, and only two polygons were permitted by each image (lung areas must be continuous). Clavicle, heart and spine regions were also annotated, but not used on the present work.

E. Model accuracy evaluation

Using lung annotations from public datasets and the ones done by a physician to imaged analog x-rays as the gold standard, the model accuracy was calculated. Intra-dataset metrics corresponds to models that were trained on and evaluated on that dataset images, whereas cross-dataset metrics correspond to models trained on one dataset and tested on different ones. As for the imaged analog dataset, all evaluations are considered as a cross-dataset evaluation, since no model was trained using that data. The MAX F1 corresponds to maximum F1-score computed, using the optimum threshold value established a priori.

V. RESULTS

A. Public digital x-ray datasets

Intra-database results show AUPRC and MAX F1 above 0.95 for all databases for both train and test samples (Tab. II, Fig. 5).

Cross-dataset tests results were also promising, with MAX F1 results above 0.9 (Tab. III), and AUPRC scores all above 0.97 (Tab. IV). Models trained with JSRT and Montgomery have similar results of MAX F1 both in intra-dataset and cross-dataset test samples (0.97 on average) with degradation of MAX F1 when challenged against Shenzhen dataset (0,96 on

¹<https://www.makesense.ai/>

TABLE II
RESULTS FROM INTRA-DATASET TRAIN AND TEST. ACCURACY METRICS SHOW PROMISING RESULTS IN BOTH TRAIN AND TEST SAMPLES.

Dataset	Montgomery	JSRT	Shenzhen
Train: MAX F1	0.992	0.990	0.969
Train: AUPRC	1.000	1.000	0.996
Validation: MAX F1	0.979	0.980	0.959
Validation: AUPRC	0.998	0.998	0.992
Test: MAX F1	0.981	0.982	0.955
Test: AUPRC	0.998	0.999	0.991

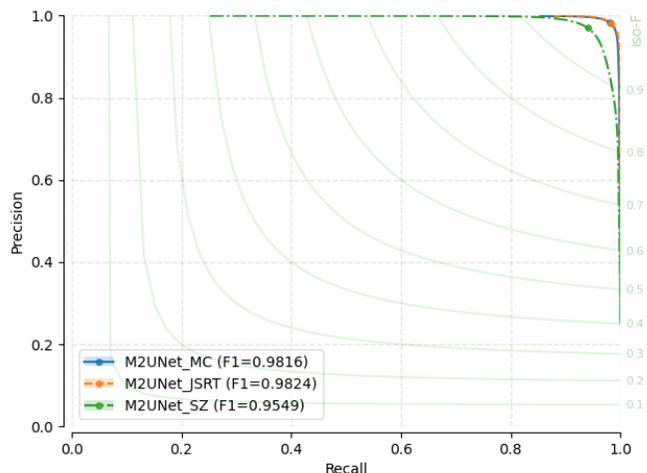


Fig. 5. Precision-Recall curves for test samples in in-database trained networks at threshold $T=0.5$. Dataset results: MC, Montgomery; JSRT, JSRT; SZ, Shenzhen

average). Shenzhen MAX F1 results tends to be more stable across all datasets (0.95 on average).

TABLE III
MAX F1 SCORES FROM CROSS-DATASET TRAIN, VALIDATION AND TEST. MONTGOMERY AND JSRT TRAINED MODELS DISPLAY SIMILAR RESULTS.

Database (sample)	Montgomery	JSRT	Shenzhen
Montgomery (train)	0.9920	0.9680	0.9560
Montgomery (validation)	0.9790	0.9670	0.9570
Montgomery (test)	0.9820	0.9700	0.9590
JSRT (train)	0.9720	0.9900	0.9630
JSRT (validation)	0.9660	0.9800	0.9620
JSRT (test)	0.9730	0.9820	0.9610
Shenzhen (train)	0.9440	0.9510	0.9690
Shenzhen (validation)	0.9440	0.9440	0.9590
Shenzhen (test)	0.9350	0.9440	0.9550

When comparing results with other more complex segmentation models trained on the same three public datasets (Tab. V), the results are found to be very similar (all F1-scores are within 0.02 of each other).

Although promising results were found analysing MAX F1 and AUPRC metrics, it was possible to verify important deviations by performing manual inspections of the predicted lung segmentations, especially on the Shenzhen dataset. In some cases, when running cross-dataset tests, images on the Shenzhen dataset that contained the patient's limbs on the x-

TABLE IV

AUPRC RESULTS FROM CROSS-DATASET TRAIN, VALIDATION AND TEST. MONTGOMERY AND JSRT TRAINED MODELS HAVE SIMILAR RESULTS.

Database (sample)	Montgomery	JSRT	Shenzhen
Montgomery (train)	1.000	0.994	0.991
Montgomery (validation)	0.998	0.995	0.990
Montgomery (test)	0.998	0.995	0.992
JSRT (train)	0.996	1.000	0.995
JSRT (validation)	0.995	0.998	0.991
JSRT (test)	0.997	0.999	0.993
Shenzhen (train)	0.986	0.988	0.996
Shenzhen (validation)	0.985	0.985	0.992
Shenzhen (test)	0.979	0.982	0.991

TABLE V

MAX F1-SCORE PERFORMANCE COMPARISON WITH OTHER EXISTING NETWORKS. CONSIDERING ONLY THE SCORES ON THE TEST PORTION OF THE DATASET

Model	Montgomery	JSRT	Shenzhen
Adv-ATTN ¹	—	0.976	—
Self-Attention DNN ²	0.982	0.971	0.960
Ours	0.981	0.982	0.955

The models correspond to: [1]Attention based Adv. UNet model [20]. [2] Self-Attention Deep Neural Network [19].

ray (a characteristic that is not present on Montgomery or JSRT datasets), incorrectly predicted the space between the arms and body as lung (Fig. 6 left).

As result of the manual inspection of the imagens with greater deviation, it was also possible to verify incorrect lung annotation especially at Shenzhen dataset. In some other cases, the inferior part of the lungs present behind radiographic shadows caused by breasts were incorrectly left out of the ground-truth annotations, where the model correctly predicted as a part of the lung (Fig. 6 center); in other images, the annotators, induced by the presence of large gastric bubbles, wrongly marked as lung the area below the diaphragm limit, where the model correctly predicted as not part of the lung (Fig. 6 right).

Another important finding is, that despite the incorrect lung segmentation (and in some cases - incorrect predictions placed even outside of the patient’s body on the x-ray), all of the trained models attained MAX F1 scores above or near 0.8. However, lungs in AP X-Rays have a established topological structure and connectivity, which is not always the case for

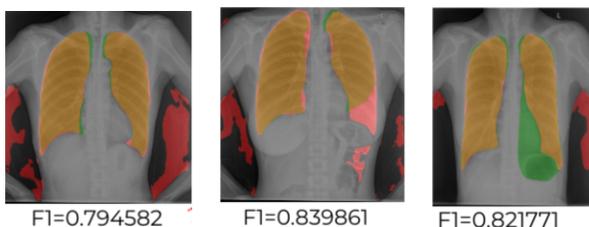


Fig. 6. Important deviations on lung segmentation were found on Shenzhen dataset. Red, model only; green, real mask; yellow, correct segmentation. Left, empty space between arm/body; center, breast shadow; right, gastric bubble



Fig. 7. Comparison of model results and physician annotated mask, images were redacted to remove personal information. Red, model only; green, real mask; yellow, correct segmentation. Left, correct segmentation for most of the lung area; center, incorrect segmentation of the neck; right, poor quality x-ray had significant reduction in accuracy.

their corresponding predicted segmentations.

The F1 score, in spite of being widely used as a metric to evaluate the performance of segmentation tasks, only measures the degree of overlap between the predicted and the ground-truth annotations, and is unable to capture how correct the shape and topological structure of the predicted lungs masks are. Given this finding, it may jeopardise and undermine AUPRC score and F1 score analysis to a certain degree, taking into account that these metrics consider each pixel independently. All this brings about a challenge to evaluate, on a large-scale, how topologically coherent the produced segmentations are. Some recent work on preserving the topology on segmentation tasks have also touched on this same challenging issue [33].

B. Imaged analogical x-rays

The trained networks were all able to correctly segment most of the lung area (Fig. 7 left), but had further degradation of their accuracy with MAX F1 at 0.78 and AUPRC at 0.9 on average (Tab. VI, Fig. 8). The models incorrectly segmented as lung part of neck, element not present on the training datasets; in some cases part of the covered negatoscope was also incorrectly segmented; in other cases, old x-rays, with poor quality or low exposure but still viable for annotation, displayed lower accuracy.

TABLE VI
RESULTS FROM ANALOG X-RAY PHOTO DATASET.

Model Dataset	MAX F1	AUPRC
Montgomery	0.808	0.931
JSRT	0.784	0.930
Shenzhen	0.775	0.841

VI. DISCUSSION AND CONCLUSION

Initial results from intra-dataset training were promising with MAX F1 and AUPRC metrics above 0.9 (Tab. II). When challenged against images from other datasets, the accuracy of the models was reduced (average MAX F1 0.96), this was

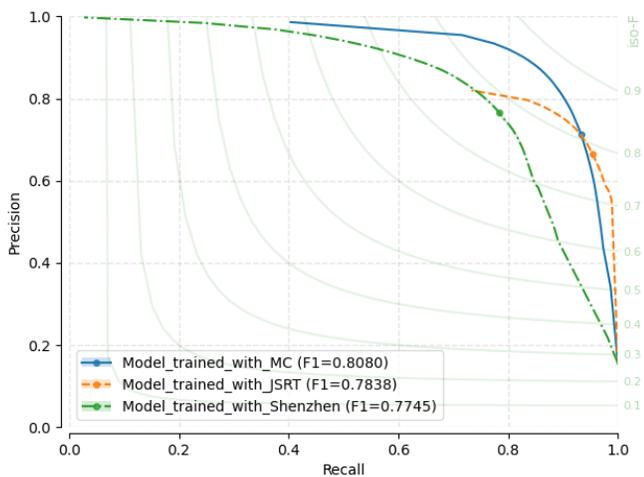


Fig. 8. Precision-Recall curves for all three network results from analog x-ray photo dataset predictions, at threshold $T=0.5$. Dataset results: MC, Montgomery; JSRT, JSRT; Shenzhen, Shenzhen

particularly more evident on the Shenzhen dataset (Tab. III). It was possible to verify with manual diagnosis that on Shenzhen datasets some of the images contained part of the patient's arms which lead the models to incorrect segment lung (Fig. 6).

When challenged against imaged analog x-rays, the models were able to correctly segment the lungs, but had its accuracy reduced even further (average MAX F1 0.78 of all three dataset-independently trained models, Tab. VI). Analog x-rays introduced new elements to classification such as neck and poor quality films (Fig. 7), both elements reduced the segmentation accuracy.

The trained models were able to correctly identify the lungs although incorrectly annotated in the training dataset, this was possible to verify by manual inspections of large deviations images. This was particularly present in Shenzhen's dataset where gastric bubble were incorrectly segmented as lungs and lungs shadowed by breasts were not segmented by human annotation (Fig. 6).

MAX F1 and AUPRC metrics although widely used in segmentation evaluation, seem not to distinguish the good from the bad in this case of study. In AP x-rays, lung area occupies an important part of the image and have a topological connectivity which at least in part breaks the assumption of independence, both important at F1 and AUPRC analysis.

This work shows preliminary results of segmentation model use to extract lungs from imaged analog x-ray and is part of a greater effort to build a CAD for tuberculosis detection to support analog equipment. A correct lung segmentation may improve detection overall accuracy of a future CAD methodology.

ACKNOWLEDGMENT

The authors would like to thank Dr. Paulo Andrade Baceti for performing lung segmentation, Diogo Coutinho Coelho

for developing the Imaging Manager App. The authors would equally like to thank the Idiap Research Institute and the State of Wallis in Switzerland for partially funding research leading to these results.

REFERENCES

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in neural information processing systems* (2012), pp. 1097–1105.
- [2] Abhishek Sharma, Daniel Raju, and Sutapa Ranjan. "Detection of pneumonia clouds in chest X-ray using image processing approach". In: *2017 Nirma University International Conference on Engineering (NUICONE)*. Nov. 2017, pp. 1–4. DOI: 10.1109/NUICONE.2017.8325607. URL: <https://ieeexplore.ieee.org/abstract/document/8325607>.
- [3] Miriam Harris, Amy Qi, Luke Jeagal, et al. "A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis". In: *PLOS ONE* 14.9 (Sept. 2019), pp. 1–19. DOI: 10.1371/journal.pone.0221339. URL: <https://doi.org/10.1371/journal.pone.0221339>.
- [4] Xiaosong Wang, Yifan Peng, Le Lu, et al. "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017). DOI: 10.1109/cvpr.2017.369. URL: <http://dx.doi.org/10.1109/CVPR.2017.369>.
- [5] Wei-jie Guan, Zheng-yi Ni, Yu Hu, et al. "Clinical Characteristics of Coronavirus Disease 2019 in China". In: *New England Journal of Medicine* (2020). DOI: 10.1056/NEJMoa2002032. URL: <https://www.nejm.org/doi/full/10.1056/nejmoa2002032>.
- [6] Ho Yuen Frank Wong, Hiu Yin Sonia Lam, Ambrose Ho-Tung Fong, et al. "Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19". In: *Radiology* 296.2 (2020). URL: <https://pubmed.ncbi.nlm.nih.gov/32216717/>.
- [7] Isabella Castiglioni, Davide Ippolito, Matteo Interlenghi, et al. "Machine learning applied on chest x-ray can aid in the diagnosis of COVID-19: a first experience from Lombardy, Italy". In: *European Radiology Experimental* 5.7 (2021). URL: <https://eurradiolexp.springeropen.com/articles/10.1186/s41747-020-00203-z#citeas>.
- [8] World Health Organization. *Chest radiography in tuberculosis detection – summary of current WHO recommendations and guidance on programmatic approaches*. 2016. ISBN: 978 92 4 151150 6. URL: <https://apps.who.int/iris/bitstream/handle/10665/252424/9789241511506-eng.pdf?sequence=1>.

- [9] World Health Organization. *Global Tuberculosis Report: Executive Summary 2020*. 2020. URL: https://cdn.who.int/media/docs/default-source/hq-tuberculosis/global-tuberculosis-report-2020/execsumm_11nov2020.pdf?sfvrsn=be12e34e_9.
- [10] Xin Li, Chengyin Li, and Dongxiao Zhu. “COVID-MobileXpert: On-Device COVID-19 Patient Triage and Follow-up using Chest X-rays”. In: (2020). arXiv: 2004.03042 [eess.IV]. URL: <https://arxiv.org/abs/2004.03042>.
- [11] Alisa I. Walz-Flannigan, Kimberly J. Brossoit, Dayne J. Magnuson, et al. “Pictorial Review of Digital Radiography Artifacts”. In: *RadioGraphics* 38.3 (2018). PMID: 29676963, pp. 833–846. DOI: 10.1148/rg.2018170038. eprint: <https://doi.org/10.1148/rg.2018170038>. URL: <https://doi.org/10.1148/rg.2018170038>.
- [12] Kai-Fabian Henning, Alexander Fritze, Eugen Gillich, et al. “Stable image acquisition for mobile image processing applications”. In: *Digital Photography XI*. Ed. by Nitin Sampat, Radka Tezaur, and Dietmar Wüller. Vol. 9404. International Society for Optics and Photonics. SPIE, 2015, pp. 177–186. URL: <https://doi.org/10.1117/12.2076146>.
- [13] Chun-Fu Yeh, Hsien-Tzu Cheng, Andy Wei, et al. “A Cascaded Learning Strategy for Robust COVID-19 Pneumonia Chest X-Ray Screening”. In: (2020). arXiv: 2004.12786 [eess.IV].
- [14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV].
- [15] Sema Candemir, Stefan Jaeger, K. Palaniappan, et al. “Lung Segmentation in Chest Radiographs Using Anatomical Atlases With Nonrigid Registration”. In: *IEEE Transactions on Medical Imaging* 33 (Feb. 2014), pp. 577–590. DOI: 10.1109/TMI.2013.2290491.
- [16] Aleksandr Zotin, Yousif Hamad, Konstantin Simonov, et al. “Lung boundary detection for chest X-ray images classification based on GLCM and probabilistic neural networks”. In: *Procedia Computer Science* 159 (2019). Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 23rd International Conference KES2019, pp. 1439–1448. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2019.09.314>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050919315145>.
- [17] Johnatan Carvalho Souza, João Otávio Bandeira Diniz, Jonnison Lima Ferreira, et al. “An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks”. In: *Computer Methods and Programs in Biomedicine* 177 (2019), pp. 285–296. ISSN: 0169-2607. DOI: <https://doi.org/10.1016/j.cmpb.2019.06.005>. URL: <https://www.sciencedirect.com/science/article/pii/S0169260719303517>.
- [18] Raghavendra Selvan, Erik B. Dam, Nicki S. Detlefsen, et al. “Lung Segmentation from Chest X-rays using Variational Data Imputation”. In: (2020). arXiv: 2005.10052 [eess.IV].
- [19] Minki Kim and Byoung-Dai Lee. “Automatic Lung Segmentation on Chest X-rays Using Self-Attention Deep Neural Network.” In: *Sensors (Basel, Switzerland) vol. 21,2* 369. 7 (2021).
- [20] Gusztáv Gaál, Balázs Maga, and András Lukács. “Attention U-Net Based Adversarial Architectures for Chest X-ray Lung Segmentation”. In: (2020). arXiv: 2003.10304 [eess.IV]. URL: <https://arxiv.org/abs/2003.10304>.
- [21] Young-Gon Kim, Kyungsang Kim, Dufan Wu, et al. “Deep Learning-based Four-region Lung Segmentation in Chest Radiography for COVID-19 Diagnosis”. In: (2020). arXiv: 2009.12610 [eess.IV].
- [22] Tim Laibacher and André Anjos. *On the Evaluation and Real-World Usage Scenarios of Deep Vessel Segmentation for Retinography*. 2019. arXiv: 1909.03856 [cs.CV]. URL: <https://arxiv.org/abs/1909.03856>.
- [23] Mark Sandler, Andrew Howard, Menglong Zhu, et al. *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. 2019. arXiv: 1801.04381 [cs.CV].
- [24] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: 1409.1556 [cs.CV].
- [25] Vladimir Iglovikov, Selim Seferbekov, Alexander Buslaev, et al. “TernausNetV2: Fully Convolutional Network for Instance Segmentation”. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2018, pp. 228–2284. DOI: 10.1109/CVPRW.2018.00042.
- [26] Liangchen Luo, Yuanhao Xiong, Yan Liu, et al. *Adaptive Gradient Methods with Dynamic Bound of Learning Rate*. 2019. arXiv: 1902.09843 [cs.LG].
- [27] Tim Laibacher and André Anjos. *On the Evaluation and Real-World Usage Scenarios of Deep Vessel Segmentation for Retinography*. 2019. arXiv: 1909.03856 [cs.CV]. URL: <https://arxiv.org/abs/1909.03856>.
- [28] Shiraishi J.; Katsuragawa S.; Ikezoe J.; Matsumoto T.; Kobayashi T.; Komatsu K.; Matsui M.; Fujita H.; Kodera Y.; Doi K. “Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists’ detection of pulmonary nodules.” In: *American Journal of Roentgenology*. 174 (2000), pp. 71–74.
- [29] Stefan Jaeger, Sema Candemir, Sameer Antani, et al. “Two public chest X-ray datasets for computer-aided screening of pulmonary diseases.” In: *Quantitative imaging in medicine and surgery*. 4.6 (2014), p. 475. ISSN: 2223-4306.
- [30] Youbao Tang, Yuxing Tang, Jing Xiao, et al. *XLsor: A Robust and Accurate Lung Segmentor on Chest X-Rays Using Criss-Cross Attention and Customized Radiorealistc Abnormalities Generation*. 2019. arXiv: 1904.09229 [cs.CV].

- [31] *Lung Masks for Shenzhen Hospital Chest X-ray Set*. <https://www.kaggle.com/yoctoman/shcxr-lung-mask>. 2018. URL: <https://www.kaggle.com/yoctoman/shcxr-lung-mask>.
- [32] Jia Deng, Wei Dong, Richard Socher, et al. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.
- [33] James Clough, Nicholas Byrne, Ilkay Oksuz, et al. “A Topological Loss Function for Deep-Learning based Image Segmentation using Persistent Homology”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021), pp. 1–1. ISSN: 1939-3539. DOI: 10.1109/tpami.2020.3013679. URL: <http://dx.doi.org/10.1109/TPAMI.2020.3013679>.