Pneumonia Detection with Deep Convolutional Architecture

Abdullah Faqih Al Mubarok
Biomedical Engineering
Institut Teknologi Bandung
Bandung, Indonesia
abdullah-faqih@students.itb.ac.id

Dominique Jeffrey A. M.
Electrical Engineering
Institut Teknologi Bandung
Bandung, Indonesia
dominiquejeffrey@students.itb.ac.id

Ahmad Habbie Thias
Biomedical Engineering
Institut Teknologi Bandung
Bandung, Indonesia
ahmad_habbie@s.itb.ac.id

***Abstract*— Pneumonia is a respiratory disease caused by an infection in the air sacs of the lungs. Patient with pneumonia will experience inflammation of the alveoli, accompanied by the presence of fluids in the air sacs. Using the intensity from thorax x-ray images, a radiologist can diagnose whether pneumonia exist or not. Computer-aided detection (CAD) can enhance the radiologist diagnostic capabilities by giving radiologist a second opinion. CAD system can be developed by several techniques including deep convolutional architecture. This paper aims to know the performance of two widely known deep convolutional architecture such as residual network and mask-RCNN in classifying and detecting pneumonia. In addition, the results will be compared and evaluated.**

Keywords—pneumonia, deep convolutional architecture, residual network, mask-RCNN.

I. Introduction

Pneumonia is an infection that inflames the air sacs in one or both lungs. The air sacs may fill with fluid or pus (purulent material), causing cough with phlegm or pus, fever, chills, and difficulty breathing. A variety of organisms, including bacteria, viruses and fungi, can cause pneumonia. This disease is very serious in toddlers and elderly patients who have a weak immune system. Pneumonia kills 920.000 toddlers worldwide in 2015 [1].

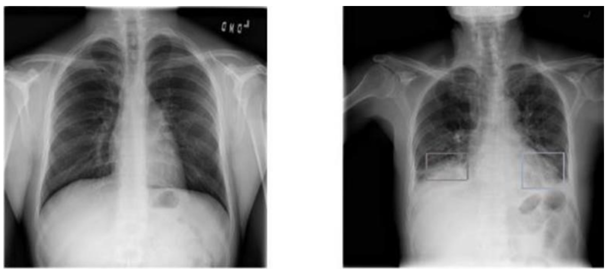


Fig. 1. **a**(left) chest radiogaph of normal and **b**(right) patient with pneumonia

Pneumonia can be detected using radiography such as conventional radiography, CT, or MRI. Doctor will examine radiograph of patient's chest to detect pneumonia. In addition, the golden standard for examining pneumonia is accompanied by the examination of patient's medical record and phlegm.

Soft tissue will produce dark color on radiograph since it can be penetrated by X-rays. Weakening X-ray intensities on hard tissues such as bones will produce bright color. Figure 1 shows that the chest cavity is clearly visible (dark color) since most of the lung cavities contain air. Whereas in patients with pneumonia, there are fluids which fill the air sacs, causing chest cavity on radiograph image looks brighter as shown in figure 1b. The brighter color on lung cavity may be caused by several abnormalities [3] such as cancer cells, swelling of blood vessels, heart abnormalities, and pneumonectomy. Figure 2 represenst a weakening of intensity in the lung cavities due to cancer.

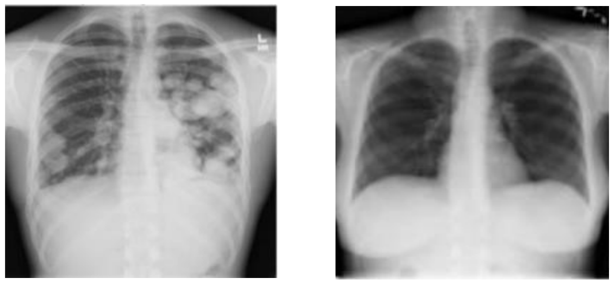


Fig. 2. **a**(left) chest abnormality caused by lung cancer and **b**(right) normal lung

In the last decade, computer-aided detection (CAD) has been developing so fast. The purpose of CAD is to assist radiologists in medical images interpretation by using second opinions from the computer systems. CAD can help to improve the diagnostic accuracy, lighten the burden of workload, and improve inter- and intra-reader variability [3]. Generally, CAD consists of the integration between image segmentation, feature extraction, and classification. Several different techniques are used for the development of CAD such as deep convolutional architecture.

Deep convolutional architecture has been led to a several breakthroughs for image classification and object detection [4, 5] since it can integrate the image information from lower and higher features. The information can be enriched with the addition of stacked layers. There are many different achitecture which widely used for image classification such as VGG16 and ResNet50. Whereas Faster R-CNN and YOLO are used for detecting object. Each of those networks had been optimized to precide the performance of previous developed architecture.

The previous studies showed that deep convolutional architecture could be used for detecting lung abnormalities from thorax x-ray images [8, 9, 10]. Indeed, a simple CNN architecture can also produce good accuracy [8]. Nevertheless, the authors of [8] didn't show the other important performance evaluation aspects such as sensitivity and specificity. Those aspects can evaluate "naive behavior" when dealing with unbalanced false positive and false negative errors. More complex architecture such as GoogLeNet and DenseNet, which are used in [9] and [10], showed good results with appropriate performance evaluation.

In this paper, we implement two architectures : deep residual network which was introduced by Microsoft researchers in ILSVRC 2015 and mask-RCNN by Facebook AI researchers. Those networks are used to develop CAD for pneumonia detection with thorax x-ray images. We also represent, compare and evaluate the performance of two architectures with simple evaluation, yet respectable.

978-1-5386-8448-1/19/$31.00 ©2019 IEEE

II. METHODOLOGY

A. Dataset

In this paper, we use Kaggle RSNA Pneumonia Detection Challenge dataset from the Radiological Society of North America. The dataset consists of 26.684 data, which are in DICOM format, divided into 21.684 training data, 2.000 validation data, 3.000 testing data. The dataset is provided as a set of patientIds, bounding boxes, and target values. In addition, The dataset is grouped into two groups, pneumonia and non-pneumonia. However, the non-pneumonia groups contain another two groups which are normal and non-pneumonia abnormal lungs.

B. Residual Network

Accuracy of the trained model will rapidly degrade if network dept increases. The problem is caused by overfitting from layer addtion[6]. Instead of hoping each feaw layers directly fit the underlying mapping, residual network uses residual maping. Consider $H(x)$ as desired underlying mapping. If we let $H(x)$ to be fitted in other mapping $F(x) = H(x) - x$, the underlying mapping is recast to $F(x) + x$ as depicted on Fig. 3.

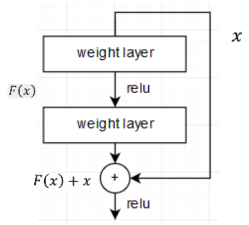
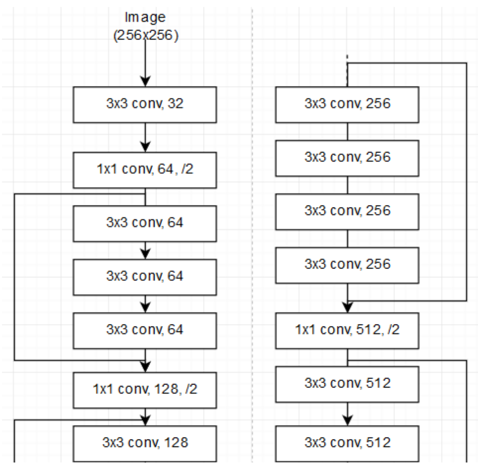


Fig. 3. Building block of residual learning

For this paper, we implement a residual network inspired from [6]. The residual mapping can be used when the input and output are in the same dimension. Batch normalization is used right after each convolution and activation. We use Adam optimizer with 32 mini batches. Cosine annealing is used for learning rate to optimize the training process. For the loss function, we combine IoU and binary cross entropy. We also implement pooling block to reduce the amount of parameters and computation in the network, hence control overfitting. Fig. 4 shows the designed residual network.



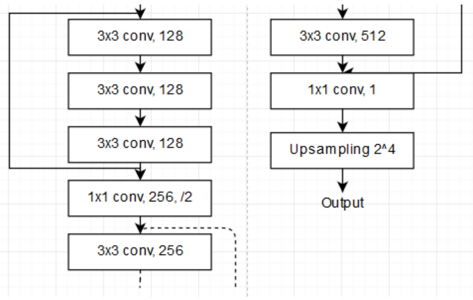


Fig. 4. Designed residual block

C. Mask Regional CNN

Mask Regional CNN (mask-RCNN) is the further development of Faster Regional CNN algorithm. This algorithm usually used in object localization and recognition in an image by combining object detection and semantic segmentation[7]. The objective of object detection is to localize each object on the image using a bounding box. Meanwhile, semantic segmentation goal is to classify each pixel into a fixed set of categories using object delineation.

Faster Regional CNN algorithm is involved in the object detection process. This algorithm consists of 2 stages, Region Proposal Network (RPN) that proposes region of interest (RoI) candidate. The second stage extracts the features using RoI-align and classifies the class of the object inside RoI.

To extract feature, our model uses the bottom-up and up-bottom extraction path. Resnet50 is involved in the bottom-up extraction path. Meanwhile, the up-bottom extraction path is using FPN (Feature Pyramid Network). Output feature of FPN then convoluted 1X1 with resnet50 output feature.

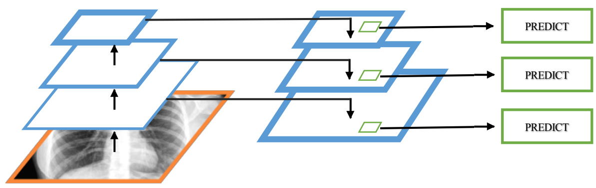


Fig. 5. Visualization of Residual Network and Feature Pyramid Network

Moreover, semantic segmentation will convolute the RoI with k^2 size kernel. This kernel will determine which part of the bounding box is an object.

These are our configuration in building the mask-RCNN model:

- 8 images per batch
- Resnet50 as Backbone
- 2 Classes, 1 for background and the other for the pneumonia
- Image dimension is 128 X 128
- Anchor sizes are 8X8, 16X16, 32X32, and 64X64
- ROIs from RPN generated are 100 per Image
- Maximum instances per image is 3
- Amount of Anchors used for RPN training per image is 200 anchor
- 125 steps per epoch
- Top down pyramid is 32

III. Result and Discussion

A. Residual Network

The Residual network provides the output with bounding boxes if the predicted images contain pneumonia. The bounding boxes are produced as a result of training the network with pneumonia masking as a target. We use confidence level greater than 0.7 so that we can get the optimal results from this network. Fig. 5 shows a random test data which contains pneumonia.

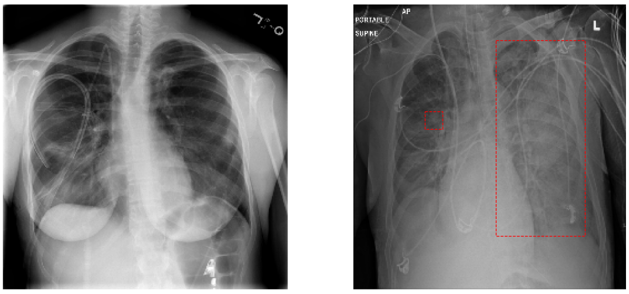


Fig. 6. Residual network random output from test data

To describe the performance of residual network, we use confusion matrix. From the confusion matrix, we can determine the accuracy, sensitivity, and specificity of an architecture given the true values. Table 1 shows the confusion matrix of the trained residual network.

TABLE I. CONFUSION MATRIX OF TRAINED RESIDUAL NETWORK WITH 0.7 CONFIDENCE LEVEL

True Label\Prediction	0	1
0	2200	320
1	120	340

From the above table, we can get the value of accuracy, sensitivity, and specificity as follow

TABLE II. PERFOMANCE OF TRAINED RESIDUAL NETWORK

Accuracy	Sensitivity	Specificity
85.60%	51.52%	94.83%

We can see that there is a contrast difference between sensitivity and specifitity from the performance of trained residual network. The difference will cause the network tend to predict a data to be a negative (non-pneumonia) class.

B. Mask Regional CNN

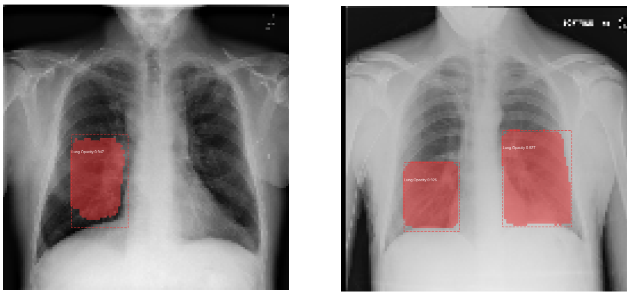


Fig. 7. Mask-RCNN network random output from test data

As we can see above, mask-RCNN gives boundary boxes and semantic areas. Images above are examples of the testing data output from trained mask-RCNN. The prediction results consist of bounding boxes, semantic segmentation of the bounding box and the confidence level of the area. The inputs of mask-RCNN are extracted features from ResNet50-FPN and RoIAllign. These inputs are trained using Fully Connected Network to do per-pixel classification. Fig. 7 shows examples of the prediction of the model.

We use 98% confidence level so that the network gives the best prediction result. We also use the confusion matrix to describe the performance of mask-RCNN.

TABLE III. CONFUSION MATRIX OF TRAINED RESIDUAL NETWORK WITH 0.98 CONFIDENCE LEVEL

True Label\Prediction	0	1
0	2177	368
1	290	164

TABLE IV. PERFOMANCE OF TRAINED MASK-RCNN

Accuracy	Sensitivity	Specificity
78.06%	36.12%	85.54%

From the table above, we also can see a big difference between true positive and true negative of the confusion matrix which leads to contrast gap between sensitivity and specificity.

C. Comparison and Analysis

TABLE V. PERFOMANCE OF TRAINED MASK-RCNN

Architecture	Accuracy	Sensitivity	Specificity
Residual Network	85.60%	51.52%	94.83%
Mask-RCNN	78.06%	36.12%	85.54%

From Table V, we can conclude that residual network performs better than mask-RCNN in all evaluation parameters. In spite of shallower and simpler architecture, residual network can generate more informative and discriminative features which lead better classification and bounding box generator. Although the authors from [7] state that mask-RCNN shows good results for Cityscapes dataset and COCO challenge dataset, The mask-RCNN may not able to extract meaningful features from thorax x-ray. We can also see that semantic segmentation from mask-RCNN gives spurious edge as we can see from Fig.6 and Fig.7.

The worse result can be caused by several factors. Firstly, the authors from [7] use COCO and Cityscapes dataset which provide clear and more vivid target object. Whereas in pneumonia detection, the pneumonia features may be difficult to be obtained by the network as a result of low quality image. Thus, the RPN alogrithm of mask-RCNN does not perform well to find which of proposed RoIs contain target object. Secondly, the configuration of mask-RCNN may not enough to get the informative and discriminative features. The authors of [7] state that more complex architecture give more good results.

Each of trained architecture also shows the contrast difference between sensitivity and specificity. The difference may caused by the unbalanced data between pneumonia and non-pneumonia classes which can cause the architecture tend to missed classify the pneumonia class. The false classification of pneumonia class is not disered for CAD since it can cause the patients tend to have normal lungs or other abnormalities.

IV. CONCLUSION

In the penumonia detection, residual network shows better performance than mask-RCNN. It may caused by the poor performance of mask-RCNN's RPN algorithm to find which of the proposed RoIs have an object since the pneumonia features are difficult to find comparing with widely used object detection dataset such as COCO. The two networks also show the contrast gap between sensitivity and spesificity which caused by unbalance dataset. In the future research, we can improve the performance of the two architectures by tuning the hyperparameters. Using more complex network structure and augmenting the unbalance dataset may also possible in the future so that we can get the best architecture for pneumonia CAD system.

REFERENCES

[1] Rui P, Kang K. National Ambulatory Medical Care Survey: 2015 Emergency Department Summary Tables.

[2] Goodman, Lawrence, "Felson's Principles of Chest Roentgenology, A Programmed Text 4th Edition", 2014.

[3] Sayed E., et. al, "Computer-aided Diagnosis of Human Brain Tumor though MRI: A Survey and a new algorithm. Expert System with Applications (41) : 2014.

[4] Krizhevsky, Alex & Sutskever, Ilya & E. Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

[5] C. Szagedy, A. Toshev, D. Erhan, "Deep neural networks for object detection", Advances in Neural Network Information Processing System (26) : 2013.

[6] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition. arXiv:1512.03385v1, 2015

[7] K. He, G. Gkioxari, P. Dollar, R. Sirshick. Mask-CNN. arXiv:1703.06870v3, 2018.

[8] Abiyev R. H., Ma'aitah M. K. S., "Deep Convolutional Neural Networks for Chest Diseases Detection", Journal of Healthcare Engineering, 2018.

[9] Cicero M. Bilbily A., Colak E., Dowdell T., Gra B., Perapaladas K., Barfett J., "Training and Validating a Deep Convolutional Neutral Network for Computer-Aided Detection and Classification of Abnormalitites on Frontal Chest Radiographs", Investigative Radiology 52(5) 281-287, 2017.

[10] Seekins J., et. al, "Deep Learning for Chest Radiograph Diagnosis: A Retrospective Comparison of The CheXNeXt Algorithm to Practicing Radiologist", PLOS Medicine 15(11), 2018