Classification and Segmentation of Pneumothorax from Chest X-rays using Deep Convolutional Neural Networks

Arman Haghanifar

September 2020

1 Introduction

Pneumothorax, also known as collapsed lung, happens when air comes through the pleural cavity between the lungs and the chest wall. Accumulation of air constantly enlarges the pleural space, resulting in compression of lung collapse and mediastinal structures. Typical symptoms of pneumothorax include sudden chest pain and shortness of breath [1]. A high-level illustration of pneumothorax is demonstrated in Fig. 1.

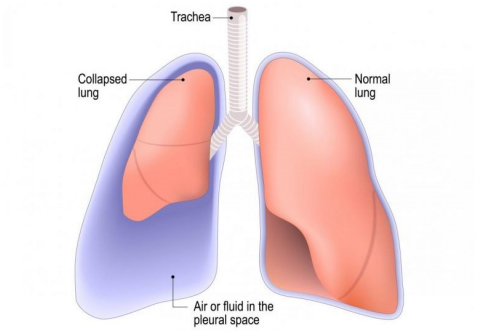
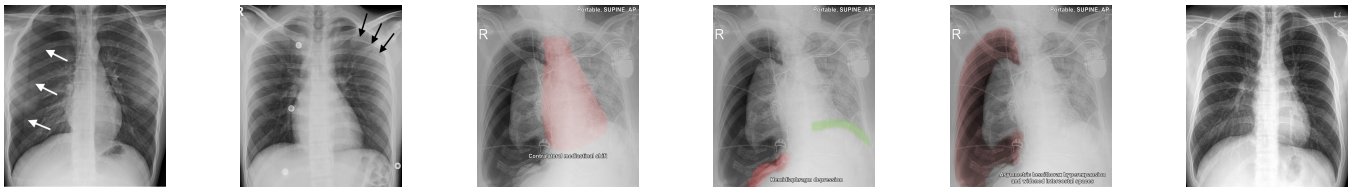


Figure 1: Illustration of pneumothorax or lung collapse

Pneumothorax can be caused by various reasons, such as other lung disease, or in some cases from an accident or injury in the chest area. However, in other cases it may occur with no specific reason [2]. Pneumothoraces caused by accidents are known as traumatic, and are more common than pneumothoraces happened without any preceding trauma [3]. The latter type is called spontaneous. Spontaneous pneumothoraces can be divided into primary and secondary. Primary spontaneous pneumothoraces occur in younger patients, whereas secondary ones occur in older patients who have underlying pulmonary diseases [4]. Pneumothoraces can also be classified as simple (no shift in mediastinal structures), tension (shift in mediastinal structures) or open (air passing through an open chest wound). Tension pneumothorax is a valve-liked fatal situation that results in air being trapped in parietal and visceral pleura inside the chest. Pleural pressure increases as more gas is coming through the pleural space. The involved lung is then pushed inside, hemidiaphragm is pushed down, mediastinal structures are displaced, and it finally results in a compression atelectasis involving the other lung. Traumatic and tension pneumothoraces are life-threatening and require immediate treatment. Therefore, early recognition of this condition is life-saving [5, 6].

2 Problem Explanation

Clinical detection of pneumothorax is complicated due to the variety of its symptoms and causes. Hence, chest x-ray (CXR) is a common tool to help radiologists diagnosing pneumothoraces. However, detection can be difficult by visual inspection, specifically when its locations are atypical or the patient has heart or lung diseases at the same time [7]. Due to its different types, pneumothorax has various manifestations when diagnosed using CXRs. A frontal CXR is usually taken from the patient for pneumothorax detection. Fig. 2 demonstrates a comparison between image features of different pneumothoraces and a normal CXR.



(a) (b) (c) (d) (e) (f)

Figure 2: Images from different pneumothorax manifestations compared to normal image

In the above figure, 2a is a right-sided secondary spontaneous pneumothorax, which can be determined with the sharp line representing the pleural edge [1]. 2b is a left-sided primary spontaneous pneumothorax, which is indicated with a subtle white pleural line [2]. 2c, 2d and 2e are different signs of a large right-sided tension pneumothorax; significant mediastinal shift to the left, depressed right hemidiaphragm and asymmetric expansion of the hemithorax, respectively [3]. 2f shows a normal CXR. Pneumothoraces are different in terms of CXR manifestations as seen above. Therefore, visual inspection by radiologist is not sufficient for efficiently detection of pneumothorax.

3 Proposed Method

Deep learning and computer vision techniques have been recently led to remarkable breakthrough in medical image analysis. Deep learning networks use large amounts of data to automatically extract data features, hence, making the training process fully automatic and resolve the need for manual feature engineering [8]. Among various deep learning algorithms, Convolutional Neural Networks (CNNs) are utilized to process images, and have achieved great success in biomedical image analysis. One of the primary CNN architectures is AlexNet [9], which hit the state-of-the-art (SOTA) in 2012 for automatic classification of general objects in ImageNet dataset. Since 2012, different architectures are introduced every year to reach the SOTA in image classification using ImageNet, such as DenseNet [10] and Xception [11]. Since introduction of EfficientNet [12] in 2019, majority of the SOTA models on ImageNet have been proposed based on the same architecture. EfficientNet is a group of CNN models which have achieved high accuracy scores with a very fewer parameters in comparison with other models. Current SOTA on ImageNet is 88.5% achieved by FixEfficientNet [13]. Recently, a number of research studies have used EfficientNet-based in medical image analysis, such as COVID-19 diagnosis in CXR images [14] or diabetic retinopathy detection using retinal fundus images [15].

[1] https://jetem.org/spontaneous_pneumothorax

[2] https://radiopaedia.org/cases/pneumothorax-38

[3] https://radiopaedia.org/cases/tension-pneumothorax-annotated-signs

Deep neural networks have also been utilized for biomedical image segmentation in recent years. In 2015, U-Net was introduced and considered as a breakthrough in using CNNs for medical image segmentation. U-Net is a Fully Convolutional Network (FCN)-based network which is constructed by multiple concatenations between convolutional blocks in encoder and decoder [16]. The proposed method in this study is to construct a U-Net network with EfficientNet-based network as its encoder. An abstract-level illustration of the proposed network is shown in Fig. 3.

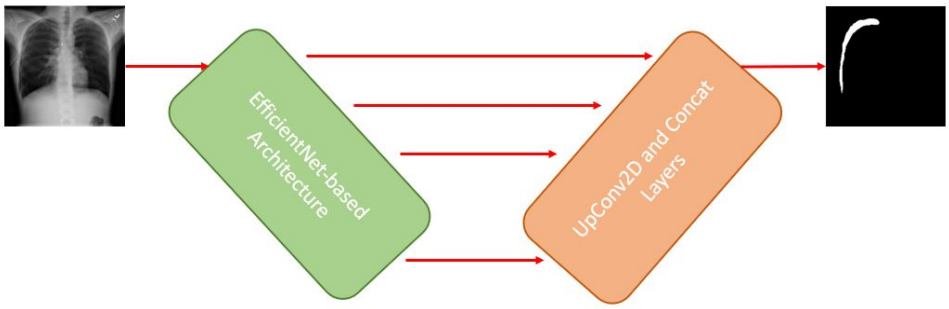


Figure 3: Illustration of proposed EfficientNet-based U-Net network

4 Materials

In this research study, CXR dataset of SIIM-ACR Pneumothorax Segmentation Challenge on Kaggle platform is used [4]. Image data is in Digital Imaging and Communications in Medicine (DICOM) format, while bounding boxes are stored in a CSV file. The DICOM format is a form of storing medical images along with patient information (metadata) in one file. All CXR images are considered as frontal, with a view position of either Anterior-Posterior (AP) or Posterior-Anterior (PA). Annotations are in the form of image IDs accompanied by Run-Length Encoding (RLE) masks. RLE is a lossless data compression method in which runs of data are saved as a single data value and count. RLE form is used for images with Pneumothorax mask values. On the other hand, images without Pneumothorax have mask value of -1. All images are in the size of 1024 $\times$ 1024 pixels.

The CXR dataset consists of 12,954 binary masks along with 12,089 images in the training-set, and 3205 images in the test-set. First step is to classify images into two categories of Pneumothorax and non-Pneumothorax. The negative images may be from healthy chests or from patients with other chest-related diseases, such as Pneumonia. In the training-set, there are more masks than the images. The reason is that some training images have multiple annotations, i.e., multiple Pneumothoraces. Second step is to segment the regions of interest; pixels which are located inside the Pneumothorax region. For the second step, 3205 images are provided which are also present in both the training-set and test-set of the first step dataset. Out of the training-set images of the first task, 9378 images are from negative class (77.57%), while 2711 are from positive one (22.43%). Hence, class imbalance is observed in the training-set for classification.

[4]https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation/

5 Model Evaluation

Typical performance metrics to evaluate a classifier model are accuracy, precision, recall and F-score. Since we have a comparison between Pneumothorax and non-Pneumothorax (may include several other diseases), we need to consider both recall and precision scores for evaluation. Besides, class imbalance may result in false high accuracy, while most correctly classified images may come from negative class. Therefore, the acceptable metric for model evaluation is F2-score; as follows.

$$F2 - score = 5 \times \frac{precision \times recall}{(4 \times precision) + recall}$$

where precision is fraction of true positive samples to all positive predicted ones, recall is fraction of true positives to all positives in the dataset. F2-score takes both false positives and false negatives into account, and is more accurate than the accuracy score when dealing with class imbalance.

For the segmentation task, two metric are considered. First, Intersection over Union (IoU), which is the area of overlap between the ground truth and the predicted segmentation, divided by the area of union between them. IoU is one of the standard metrics for segmentation model evaluation, and is computed as follows.

$$IoU(P_{true}, P_{predicted}) = \frac{P_{true} \cap P_{predicted}}{P_{true} \cup P_{predicted}}$$

Moreover, to be able to compare the results with the leaderboard, competition's official evaluation metric should also be considered. Sorenson-Dice (DSC) is an evaluation metric used to compare a predicted segmentation and its corresponding ground truth, pixel-wisely. DSC is computed as follows.

$$DSC(X, Y) = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

where X is the predicted set of pixels and Y is the ground truth. To date, highest obtained mean DSC scores are 0.9304 and 0.8679 in public and private leaderboards, respectively [5].

[5] https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation/leaderboard

References

[1] P. Zarogoulidis, I. Kioumis, G. Pitsiou, K. Porpodis, S. Lampaki, A. Papaiwannou, N. Katsikogiannis, B. Zaric, P. Branislav, N. Secen *et al.*, "Pneumothorax: from definition to diagnosis and treatment," *Journal of thoracic disease*, vol. 6, no. Suppl 4, p. S372, 2014.

[2] C. Strange, "Pleural complications in the intensive care unit," *Clinics in chest medicine*, vol. 20, no. 2, pp. 317–327, 1999.

[3] D. Gupta, A. Hansell, T. Nichols, T. Duong, J. G. Ayres, and D. Strachan, "Epidemiology of pneumothorax in england," *Thorax*, vol. 55, no. 8, pp. 666–671, 2000.

[4] T. W. Shields, *General thoracic surgery*. Lippincott Williams & Wilkins, 2005, vol. 1.

[5] R. Diaz and D. Heller, "Barotrauma and mechanical ventilation," *Stat-Pearls, Treasure Island*, 2020.

[6] A. Sharma and P. Jindal, "Principles of diagnosis and management of traumatic pneumothorax," *Journal of Emergencies, Trauma, and Shock*, vol. 1, no. 1, pp. 34–41, 2008.

[7] K.-Y. Chen, J.-S. Jerng, W.-Y. Liao, L.-W. Ding, L.-C. Kuo, J.-Y. Wang, and P.-C. Yang, "Pneumothorax in the icu: patient outcomes and prognostic factors," *Chest*, vol. 122, no. 2, pp. 678–683, 2002.

[8] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[9] 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.

[10] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "Densenet: Implementing efficient convnet descriptor pyramids," *arXiv preprint arXiv:1404.1869*, 2014.

[11] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[12] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.

[13] H. Touvron, A. Vedaldi, M. Douze, and H. Jégou, "Fixing the train-test resolution discrepancy: Fixefficientnet," *arXiv preprint arXiv:2003.08237*, 2020.

[14] G. Marques, D. Agarwal, and I. de la Torre Díez, "Automated medical diagnosis of covid-19 through efficientnet convolutional neural network," *Applied Soft Computing*, p. 106691, 2020.

[15] M. Chetoui and M. A. Akhloufi, "Explainable diabetic retinopathy using efficientnet," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 1966–1969.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.