# 2ST-UNet: 2-Stage Training Model using U-Net for Pneumothorax Segmentation in Chest X-Rays

Ayat Abedalla*, Malak Abdullah*, Mahmoud Al-Ayyoub*† and Elhadj Benkhelifa‡
*Jordan University of Science and Technology, Irbid, Jordan
†University of Manchester, Manchester, UK
‡Staffordshire University, Stoke-on-Trent, UK
ayatabedalla@gmail.com, mabdullah@just.edu.jo, mahmoud.alshboul@manchester.ac.uk, E.Benkhelifa@staffs.ac.uk

***Abstract*—Pneumothorax, also called a collapsed lung, is the presence of the air outside of the lung in the space between the lung and chest wall. It is generally diagnosed using a chest X-ray. However, for some cases, the diagnosis can be difficult as other medical conditions appear similarly. Machine Learning algorithms have been providing great assistance in detecting and locating pneumothorax lately. In this paper, we propose a 2-Stage Training system to segment images with pneumothorax. This system has been built based on U-Net, the state-of-the-art Fully Convolutional Network (FCN) architecture, with a backbone Residual Networks (ResNet-34) that is pre-trained on the ImageNet dataset. In the beginning, we train the network at a lower resolution. Then, we load the trained model weights to retrain the network with a higher resolution. Moreover, we utilize different techniques including Stochastic Weight Averaging (SWA), data augmentation, and Test-Time Augmentation (TTA). We use the chest X-ray dataset that is provided by the 2019 SIIM-ACR Pneumothorax Segmentation Challenge, which contains 12047 training images and 3205 testing images. Our experiments show that 2-Stage Training leads to better and faster network convergence. Our method achieves 0.8356 mean Dice coefficient placing it among the top 9% of competitors with a rank of 124 out of 1475.**

***Index Terms*—Pneumothorax Segmentation, 2-Stage Training, U-Net, ResNet-34, Chest X-ray, Transfer Learning, Data Augmentation, Test-Time Augmentation.**

## I. Introduction

Pneumothorax (Collapsed Lung) is the presence of air in the pleural cavity between the lungs and the chest wall. The pressure of this air causes the lung to collapse on itself [1]. Pneumothorax can be caused by a variety of reasons, such as lung diseases or defects. In some cases, it can occur from an accident or injury in the chest area, while, in other cases, it may occur without any apparent reason. Pneumothorax can lead to a life-threatening medical emergency if it is not treated immediately (subsequent dyspnea and alveoli explosion due to the presence of air) [2]. Detecting pneumothorax is complicated due to the variety of its symptoms and causes. It is usually detected by a radiologist using chest X-rays. However, it can be difficult to diagnose especially when its locations are atypical or when the patient has heart or lung diseases [3].

Machine Learning (ML) techniques have led to remarkable advances in medical image analysis. A branch of ML known as Deep Learning (DL) has been used to process massive amounts of data in order to automatically learn their features without the need for manual feature extraction/engineering [4]. One of the modern DL models is known as Convolutional Neural Networks (CNN) [5], which achieved great success in general and biomedical image analysis [6].

Image segmentation techniques, which provide a "better understanding" of the objects in the image, are extremely useful for several tasks. For medical image processing, there is a need to localize and segment objects or regions such as brain tumors [7], [8], brain tissues [9], and abdominal aortic aneurysms [10]. Recently, Fully Convolutional Networks (FCN) [11] achieved great success in medical image segmentation. This success has been mainly associated with efficient architectures such as SegNet [12] and U-Net [13].

In this study, we propose a 2-Stage Training method to segment pneumothorax in chest X-rays. The dataset we use was provided by the SIIM-ACR Pneumothorax Segmentation Challenge.[1] The proposed segmentation model has a U-Net [13] architecture with 34-layer Residual Network (ResNet-34) [14] pre-trained on the ImageNet [15] as a backbone. We have resized the original $1024 \times 1024$ X-ray images to $256 \times 256$ and $512 \times 512$ image sizes. In stage-1, the model is trained on the lower resolution of $256 \times 256$. Then, the weights of the model are loaded in stage-2 to train the model on the higher resolution of $512 \times 512$. As a result, the model achieves 0.8356 mean Dice coefficient.

This paper is organized as follows. Section II presents the related work on pneumothorax research and the top teams for the SIIM-ACR Pneumothorax Segmentation Challenge. Section III describes the dataset that are used in this paper. Section IV provides an overview of our methodology and the architecture of our approach. Section V presents experimental settings and metrics for evaluations. The results and discussion are given in Section VI. Finally, the conclusion and future work are provided in Section VII.

## II. Related Work

Segmenting and detecting pneumothorax have attracted the attention of many researchers. In [16], the authors evaluated the performance of three DL techniques: CNN, FCN, and Multiple-Instance Learning (MIL) [17], to detect and localize pneumothorax in chest X-ray images. They found that CNN had the best performance in terms of Area Under The Curve

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

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

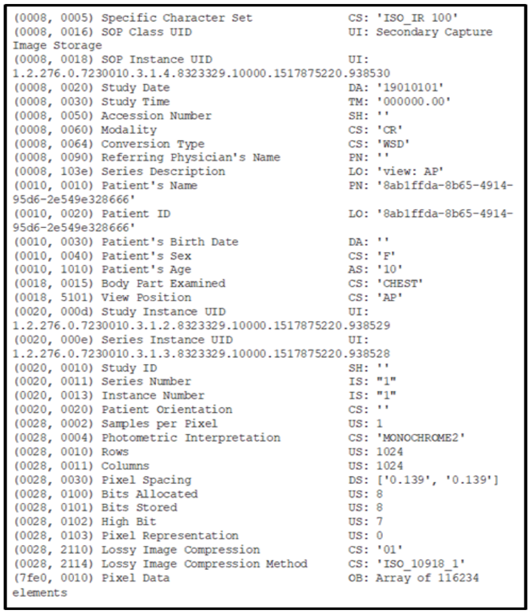
(AUC), while the other models detect the location of pneumothorax more accurately. In [18], the researchers used a deep CNN to identify moderate and large pneumothorax in chest X-rays images.

The SIIM-ACR Pneumothorax Segmentation Challenge started on August 21, 2019, as a 2-round competition. In round-1, each participating team would develop a model to segment pneumothorax based on the provided training set. The number of participants in round-1 was 1475 teams. In round-2, each team has to use the model they developed in round-1 and retrain it on an updated training data set. Then, they have to use the retrained models for inferring and predicting an unseen test data set.

The models of the top ten winning teams[2] rely mostly on deep CNN. The first-place team, Team [dsmlkz] sneddy, used U-Net model with ResNet-34, ResNet-50, and 50-layer ResNext [19] with the Squeeze-and-Excitation (SE) [20] module (SE-ResNext-50). The models were trained several times. First, these models (except for ResNet-50) were trained on size $512 \times 512$ then uptrained on size $1024 \times 1024$ with a frozen encoder in early epochs. The best model from the previous step was uptrained until convergence with a 1e-5 learning rate and 0.6 sample rate. Then, they repeated the previous step with a 0.4 sample rate. Finally, the models were uptrained with a small learning rate of 1e-6 and a sample rate of 0.5. They also used aggressive augmentation for training and applied horizontal flip TTA and triplet threshold for post-processing. The second-place team, Team X5, used classification and segmentation models in their approach. For classification, they used stacking ensemble of multi-task U-Net model with SE-ResNeXt-50, SE-ResNeXt-101 and EfficientNet-b3 [21] backbones. For segmentation, they used an averaged ensemble of U-Net model with SE-ResNeXt-50, SE-ResNeXt-101, EfficientNet-b3, EfficientNet-b5 backbones and DeepLabv3 [22] with ResNeXt-50 backbone.

## III. Dataset

In this study, we have used the chest X-ray dataset from the SIIM-ACR Pneumothorax Segmentation Challenge. The images of this dataset were released in the Digital Imaging and Communications in Medicine (DICOM)[3] format, while the annotations in the form of image IDs and Run-Length Encoding (RLE)[4] masks. The DICOM format can store medical images (pixel data) along with patient information (metadata) in one file. Figure 1 provides an example of data elements in the DICOM file. RLE is a form of lossless data compression in which runs of data are stored as a single data value and count. The relative form of RLE is used for images with pneumothorax as a mask value where pixel locations are measured from the previous end of run as shown in Figure 2. On the other hand, images without pneumothorax have a mask value of -1. All images are in size of $1024 \times 1024$ pixels. Figure 3 shows some X-ray images that contain pneumothorax.



```
(0008, 0005) Specific Character Set            CS: 'ISO_IR 100'
(0008, 0016) SOP Class UID                     UI: Secondary Capture
Image Storage
(0008, 0018) SOP Instance UID                  UI:
1.2.276.0.7230010.3.1.4.8323329.10000.1517875220.938530
(0008, 0020) Study Date                        DA: '19010101'
(0008, 0030) Study Time                        TM: '000000.00'
(0008, 0050) Accession Number                  SH: ''
(0008, 0060) Modality                          CS: 'CR'
(0008, 0064) Conversion Type                   CS: 'WSD'
(0008, 0090) Referring Physician's Name        PN: ''
(0008, 103e) Series Description                LO: 'view: AP'
(0010, 0010) Patient's Name                    PN: '8ab1ffda-8b65-4914-
95d6-2e549e328666'
(0010, 0020) Patient ID                        LO: '8ab1ffda-8b65-4914-
95d6-2e549e328666'
(0010, 0030) Patient's Birth Date              DA: ''
(0010, 0040) Patient's Sex                     CS: 'F'
(0010, 1010) Patient's Age                     AS: '10'
(0018, 0015) Body Part Examined                CS: 'CHEST'
(0018, 5101) View Position                     CS: 'AP'
(0020, 000d) Study Instance UID                UI:
1.2.276.0.7230010.3.1.2.8323329.10000.1517875220.938529
(0020, 000e) Series Instance UID               UI:
1.2.276.0.7230010.3.1.3.8323329.10000.1517875220.938528
(0020, 0010) Study ID                          SH: ''
(0020, 0011) Series Number                     IS: "1"
(0020, 0013) Instance Number                   IS: "1"
(0020, 0020) Patient Orientation               CS: ''
(0028, 0002) Samples per Pixel                 US: 1
(0028, 0004) Photometric Interpretation        CS: 'MONOCHROME2'
(0028, 0010) Rows                              US: 1024
(0028, 0011) Columns                           US: 1024
(0028, 0030) Pixel Spacing                     DS: ['0.139', '0.139']
(0028, 0100) Bits Allocated                    US: 8
(0028, 0101) Bits Stored                       US: 8
(0028, 0102) High Bit                          US: 7
(0028, 0103) Pixel Representation              US: 0
(0028, 2110) Lossy Image Compression           CS: '01'
(0028, 2114) Lossy Image Compression Method    CS: 'ISO_10918_1'
(7fe0, 0010) Pixel Data                        OB: Array of 116234
elements
```

Fig. 1. Data elements in DICOM file

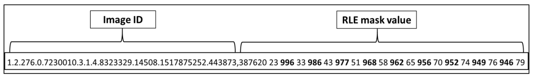


Fig. 2. Example of RLE

As mentioned earlier, this challenge had two rounds. In round-1, the training set consists of 10675 chest radiographs and 11582 masks, while the test set consist of 1372 chest radiographs. In round-2, the updated training set contains round-1 training and testing sets with updated labels, and a new testing set with unseen labels consisting of 3205 chest radiographs. In the training set for both rounds, there are more masks than the images. This indicates that some training images have multiple annotations, i.e., multiple regions of interest.

The round-1 training set contains 2379 positive cases of finding pneumothorax, while round-2 training set has 2669 positive cases of finding pneumothorax out of 12047 cases. Table I shows more details for the dataset.

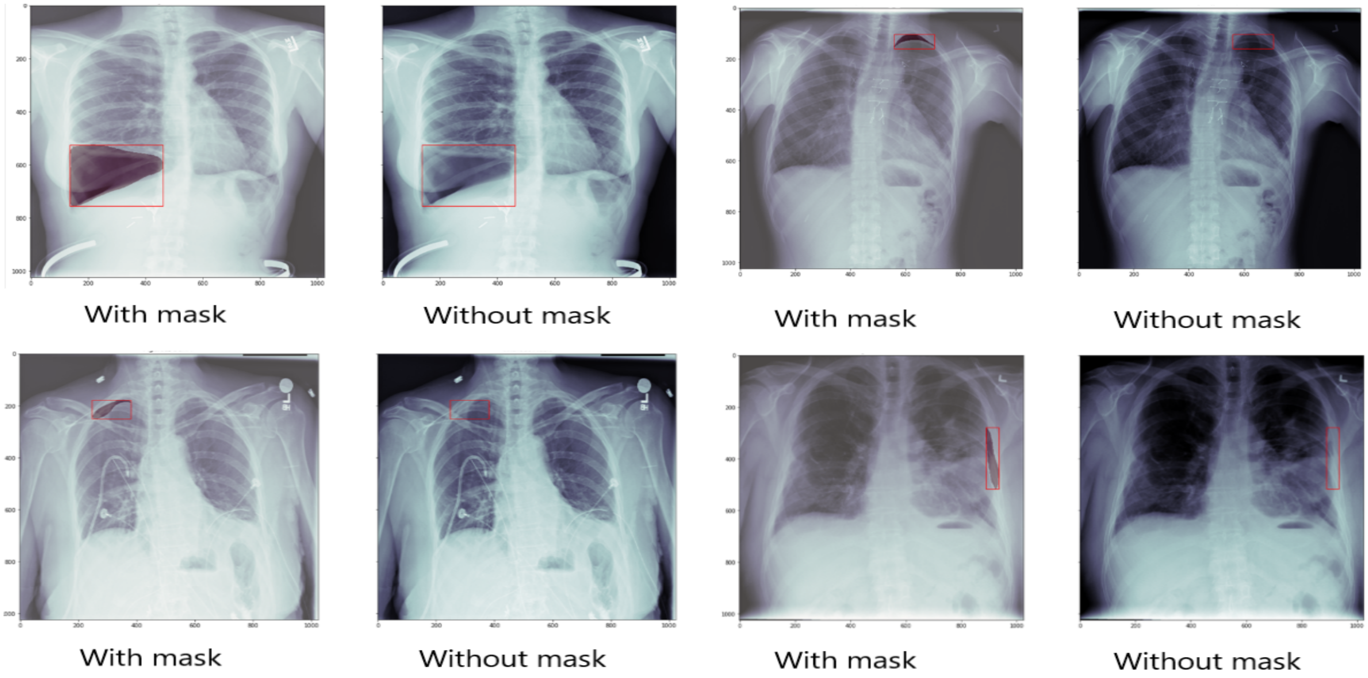
TABLE I
Dataset overview

| Attribute | Round-1 | Round-2 |
|---|---|---|
| Number of training samples | 10675 | 12047 |
| Number of test samples | 1372 | 3205 |
| Number of positive samples | 2379 | 2669 |
| Number of RLE masks | 11582 | 12954 |

[2] The top 10 winning teams models of the Pneumothorax Challenge are available on https://siim.org/page/pneumothorax_challenge

[3] https://www.dicomstandard.org/current/

[4] https://en.wikipedia.org/wiki/Run-length_encoding



With mask

Without mask

With mask

Without mask

With mask

Without mask

With mask

Without mask

Fig. 3. Samples of chest X-ray image containing pneumothorax

## IV. Proposed Approach

In this section, we provide details about the image pre-processing and augmentation techniques that have been used in our approach. Then, we present the segmentation network, training, and testing details. The detailed methodology of our approach is illustrated in Figure 4.

### A. Data Pre-processing and Augmentation

First, we extract the pixel array in a resolution of $1024 \times 1024$ pixels from the DICOM files and convert the grayscale images (that have only 1 channel) to RGB colored images (with three channels). Then, we resize these images to $256 \times 256$ pixels and $512 \times 512$ pixels for the 2-Stage Training steps. Data augmentation is a useful way to reduce the generalization error (overfitting) of models by increasing the amount of training data and adding a variety of distortions and noise to the training data in order to strengthen and improve the robustness of the model [23]. We have applied four groups of data augmentation techniques including horizontal flip, one of random contrast, random gamma, and random brightness, one of elastic transform, grid distortion, and optical distortion, and random sized crop. Finally, we have normalized the pixel values so the range of pixel intensity values are between 0 and 1.

### B. Model Architecture

In our approach, we use U-Net [13] as a segmentation model, which is an extension of FCN. The backbone used is ResNet-34 [14] pre-trained on the ImageNet [15] dataset. This is an example of transfer learning, which is an effective technique where a model that is trained for one problem can be reused as initialization for another model that is to be trained on another similar problem [24]. The network architecture is based on a 2-Stage Training technique. First, we have trained our network on lower resolution images ($256 \times 256$). Then, the model's weights are loaded and the learning rate is initialized to train the model on higher resolution ($512 \times 512$).

Our ResNet34-UNet segmentation network has an encoder and a decoder. The encoder is built by removing the fully connected layer and the global average pooling layer from the end of ResNet-34. The decoder has five blocks, each consisting of an up-sampling layer followed by repeated two times of convolution layer, batch normalization layer, and ReLU activation layer. In the first four blocks of the decoder, the feature maps after up-sampling are concatenated with the feature maps from the same sized part in the encoder. Finally, we apply a convolution layer followed by sigmoid activation to output the binary masks.

### C. Training and Testing Details

We train our network for 100 epochs with batch size 64 on lower resolution ($256 \times 256$) images. Then, the model weights are loaded and the learning rate is initialized to train the model for 70 epochs with batch size 16 on the $512 \times 512$ images. We use the Adam [25] optimizer with an initial learning rate of 0.001, which is relatively dropped per epoch using the cosine annealing learning rate technique [26]. Also, we apply Stochastic Weight Averaging (SWA) [27] to converge more quickly to the wider optima that enables the model to generalize well. The SWA setup for the $256 \times 256$ images has been given for the last five epochs, while for the $512 \times 512$ images, it is for the last three epochs. As a loss function, we

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

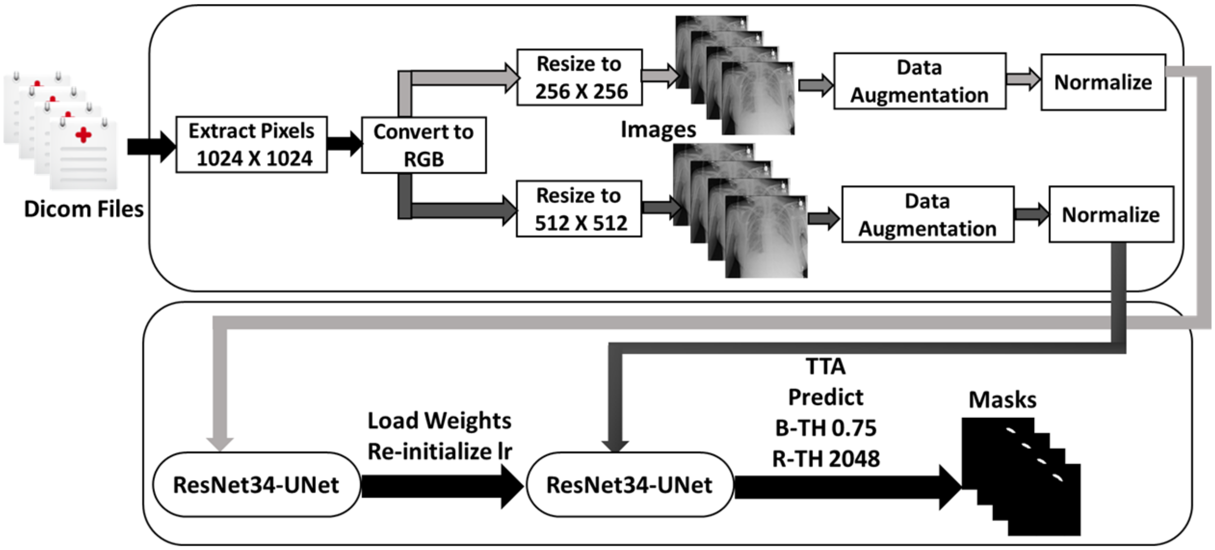


Fig. 4. Our methodology for pneumothorax segmentation

use a combination of Binary Cross-Entropy (BCE) and Dice loss. BCE is defined as follows.

$$\text{BCE} = -1/N \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \quad (1)$$

where $N$ is the number of training samples, $y$ is the ground truth value, and $p$ is the predicted value. The Dice loss is defined as follows.

$$\text{DSCL} = 1 - \frac{2 \times |\text{X} \bigcap \text{Y}|}{|\text{X}| + |\text{Y}|} \quad (2)$$

where $X$ is the predicted set of pixels and $Y$ is the ground truth. The expression for the loss function is obtained as follows.

$$\text{BCE-Dice Loss} = \text{BCE} + \text{DSCL} \quad (3)$$

For testing, the images have been resized to 512 × 512, and converted to RGB colors with normalization. As for the post-processing steps, we apply horizontal flip, Test-Time Augmentation (TTA) [28], to test images. Eventually, the final prediction is the average prediction for all images. Moreover, we use the Removal Threshold (R-TH) for each predicted mask to reduce false positives. The components that are less than the minimum size of pixels are removed.

# V. Experiments

## A. Experiments details

We conduct several experiments to evaluate our model as shown in Table II. These experiments are tuned based on the validation data. We take from each experiment the best one to evaluate the model on round-1 test set. Finally, the best experiment is used to infer round-2 test set.

For the first experiment (Exp1), we train the network at a resolution of 256×256 and 1 channel for 50 epochs with batch size 40. To enable us to train the network of grayscale images with pre-trained weights, we add the convolution layer to map 1-channel data to 3-channels. We use the Adam optimizer and ReduceLROnPlateau[5] scheduler to reduce the learning rate by a factor of 0.2 if no improvement is seen for 5 patience based on the validation loss. The Binarization Threshold (B-TH) is 0.5 for the prediction.

As for the second experiment (Exp2), we train the segmentation network on the 256 × 256 resoultion images with 3 channels for 35 epochs with 40 batch size. Moreover, the B-TH is 0.55 for the prediction.

In the third experiment (Exp3), we load the weights from the Exp2 model and reinitialize the learning rate. Then, we train the network on the 512 × 512 resolution images with 3 channels for 10 epochs with batch size 14. The B-TH is 0.55 for the prediction.

For the fourth experiment (Exp4), we first train our network on the 256 × 256 resolution images, and 3 channels for 60 epochs with batch size 40. Then, we load weights and reinitialize the learning rate to retrain the network on the 512 × 512 resolution images and 3 channels for 29 epochs with batch size 14. The B-TH is 0.75 for the prediction.

Finally, the configuration of the fifth experiment (Exp5) is the same as Exp4's, except that we train the model on the 256 × 256 resolution images for 100 epochs, and 512 × 512 for 70 epochs.

We divide the data into 80% for training and 20% for validation in the first experiment and 90% for training and 10% for validation in the other experiments. All experiments have been trained using the U-Net model with the ResNet-34 backbone. We use SWA with cosine annealing learning rate techniques for all experiments except the first one. As for the prediction, R-TH is 2048, which means that we remove images with a total mask size smaller than 2048 pixels.

[5] https://keras.io/callbacks/\#reducelronplateau

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

TABLE II
CONDUCTED EXPERIMENTS AND THEIR SETTINGS

| Experiment | SWA | B-TH | Channels | 256 × 256 | | 512 × 512 | |
|---|---|---|---|---|---|---|---|
| | | | | Batch size | Epochs | Batch size | Epochs |
| Exp1 | ✗ | 0.5 | 1 | 40 | 50 | N/A | N/A |
| Exp2 | ✓ | 0.55 | 3 | 40 | 35 | N/A | N/A |
| Exp3 | ✓ | 0.55 | 3 | 40 | 35 | 14 | 10 |
| Exp4 | ✓ | 0.75 | 3 | 40 | 60 | 14 | 29 |
| Exp5 | ✓ | 0.75 | 3 | 64 | 100 | 16 | 70 |

### B. Environment Settings

All experiments code is written in Python 3. The first four experiments are run on the Colaboratory cloud service provided by Google,[6] whereas the last experiment is run on a local server with 32 GB of RAM and Nvidia Titan Xp GPU. For Exp5 (2-Stage Training), each training epoch takes about 4 minutes in stage-1 and 8 minutes in stage-2.

### C. Metric

During training, we have used Intersection over Union (IoU) [29] as a metric. IoU is the area of overlap between the ground truth ($P_{\text{true}}$) and the predicted segmentation ($P_{\text{predicted}}$) divided by the area of union between them. The following formula computes the IoU.

$$\text{IoU}(P_{\text{true}}, P_{\text{predicted}}) = \frac{P_{\text{true}} \bigcap P_{\text{predicted}}}{P_{\text{true}} \bigcup P_{\text{predicted}}} \quad (4)$$

The competition's official evaluation metric is the mean Dice similarity coefficient (DSC) [30]. This metric is used to compare the pixel-wise agreement between a predicted segmentation and its corresponding ground truth. The following formula computes the Dice coefficient.

$$\text{DSC}(X, Y) = \frac{2 \times |X \bigcap Y|}{|X| + |Y|} \quad (5)$$

where $X$ is the predicted set of pixels and $Y$ is the ground truth.

## VI. RESULTS AND DISCUSSION

During round-1 of the competition, the score results reflected the entire round-1 test set, whereas, during round-2 of the competition, the score was calculated with only 1% of round-2 test set. Upon the completion of the competition, scores on the private leaderboard were calculated with the remaining 99% of round-2 test data.

Table III shows the results of the IoU score for all experiments on the validation set with and without using the TTA technique. The results indicate that the use of TTA is highly effective and leads to better predictive performance. We choose the best experiment from each experiment on the validation set to predict round-1 test set. Table IV shows the results of the Dice score for round-1 test set. The 2-Stage Training model achieves a Dice score of 0.8502, and this indicates that 2-Stage Training improves the learning process of the model as it learns the features of its training at different resolutions.

TABLE III
IOU SCORE FOR VALIDATION SET

| Experiment | IoU Score | |
|---|---|---|
| | Without TTA | Using TTA |
| Exp1 | 0.7242 | 0.7395 |
| Exp2 | 0.7349 | 0.7523 |
| Exp3 | 0.7105 | 0.7401 |
| Exp4 | 0.7422 | 0.7631 |
| Exp5 | 0.7638 | 0.7822 |

TABLE IV
DICE SCORE FOR ROUND-1 TEST SET

| Experiment | Dice Score |
|---|---|
| Exp1 | 0.8125 |
| Exp2 | 0.8154 |
| Exp3 | 0.8300 |
| Exp4 | 0.8407 |
| Exp5 | **0.8502** |

For round-2 of the competition, a new test was released and the training set was updated to include round-1 test set. We choose the best experiment (Exp5) 2-Stage Training to infer round-2 test set. This experiment achieves the highest score on both the validation set and round-1 test set using the TTA technique. As shown in Table V, we infer round-2 test set using the model that has been trained on round-1 training set. Then, we train the model on round-2 training set to predict round-2 test set. We also make a comparison using TTA and without using it to confirm the effectiveness of our approach. The results indicate that the use of TTA technique improves the prediction, where the horizontal flip transformation of the test images increases the chances of capturing the target shape and predicting performance. Our model achieves a Dice score of 0.8356. Table VI shows the performance of our method compared to the winning methods of the Pneumothorax Challenge on the test dataset. We notice that, for this task, the high resolution images increases the performance in addition to the use of deeper and more powerful backbones. It is worth mentioning that we have been awarded a bronze medal on Kaggle for being in the top 9% of competitors with a rank of 124 out of 1475.

## VII. CONCLUSION

In this paper, a 2-Stage Training system is proposed to segment pneumothorax based on an elegant state-of-the-art architecture called U-Net. The backbone used is ResNet-34 pre-trained on the ImageNet dataset and the dataset consists

[6] https://research.google.com/colaboratory/faq.html

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

TABLE V
ROUND-2 RESULTS OF 2ST-UNET SYSTEM WITH DICE SCORE

| Train set | TTA | Dice Score | |
|---|---|---|---|
| | | Public Leaderboard | Private Leaderboard |
| Round-1 Train | ✗ | 0.9005 | 0.8247 |
| Round-1 Train | ✓ | 0.8938 | 0.8330 |
| Round-2 Train | ✗ | 0.9011 | 0.8298 |
| Round-2 Train | ✓ | **0.9023** | **0.8356** |

TABLE VI
DICE SCORE COMPARED WITH THE TOP WINNING METHODS ON THE PNEUMOTHORAX ROUND-2 TEST SET

| Team | Model | Image size | Dice Score |
|---|---|---|---|
| [dsmlkz] sneddy | U-Net | 1024 × 1024 | 0.8679 |
| X5 | Deeplabv3+, U-Net | 1024 × 1024 | 0.8665 |
| 2ST-UNet | U-Net | 512 × 512 | 0.8356 |

of chest X-ray images provided by the 2019 SIIM-ACR Pneumothorax Segmentation Challenge. We have employed data augmentation, SWA and TTA techniques to improve the network's predictions. Our method achieves a mean Dice coefficient of 0.8356 and was ranked 124 out of 1475 competitors.

## ACKNOWLEDGEMENT

We gratefully acknowledge the support of the Deanship of Research at the Jordan University of Science and Technology for supporting this work via Grant #20190180 in addition to NVIDIA Corporation for the donation of the Titan Xp GPU used for this research.

## 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] 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.

[4] G. Al-Bdour, R. Al-Qurran, M. Al-Ayyoub, and A. Shatnawi, "A detailed comparative study of open source deep learning frameworks," *arXiv preprint arXiv:1903.00102*, 2019.

[5] S. Haykin and B. Kosko, *GradientBased Learning Applied to Document Recognition*. IEEE, 2001. [Online]. Available: https://ieeexplore.ieee.org/document/5265772

[6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[7] A. Kermi, I. Mahmoudi, and M. T. Khadir, "Deep convolutional neural networks using u-net for automatic brain tumor segmentation in multimodal mri volumes," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 37–48.

[8] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[9] Z. Kong, T. Li, J. Luo, and S. Xu, "Automatic tissue image segmentation based on image processing and deep learning," *Journal of healthcare engineering*, vol. 2019, 2019.

[10] M. de Bruijne, B. van Ginneken, M. A. Viergever, and W. J. Niessen, "Interactive segmentation of abdominal aortic aneurysms in cta images," *Medical Image Analysis*, vol. 8, no. 2, pp. 127–138, 2004.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[13] 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.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[16] A. Gooßen, H. Deshpande, T. Harder, E. Schwab, I. Baltruschat, T. Mabotuwana, N. Cross, and A. Saalbach, "Deep learning for pneumothorax detection and localization in chest radiographs," *arXiv preprint arXiv:1907.07324*, 2019.

[17] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.

[18] A. G. Taylor, C. Mielke, and J. Mongan, "Automated detection of moderate and large pneumothorax on frontal chest x-rays using deep convolutional neural networks: A retrospective study," *PLoS medicine*, vol. 15, no. 11, p. e1002697, 2018.

[19] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

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

[22] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[23] M. Ebrahim, M. Alsmirat, and M. Al-Ayyoub, "Performance study of augmentation techniques for hep2 cnn classification," in *2018 9th International Conference on Information and Communication Systems (ICICS)*. IEEE, 2018, pp. 163–168.

[24] M. Ebrahim, M. Al-Ayyoub, and M. A. Alsmirat, "Will transfer learning enhance imagenet classification accuracy using imagenet-pretrained models?" in *2019 10th International Conference on Information and Communication Systems (ICICS)*. IEEE, 2019, pp. 211–216.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[26] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.

[27] P. Izmailov, D. Podoprikhin, T. Garipov, D. Vetrov, and A. G. Wilson, "Averaging weights leads to wider optima and better generalization," *arXiv preprint arXiv:1803.05407*, 2018.

[28] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren, "Test-time augmentation with uncertainty estimation for deep learning-based medical image segmentation," *arXiv preprint arXiv:1807.07356*, 2018.

[29] P. Jaccard, "The distribution of the flora in the alpine zone. 1," *New phytologist*, vol. 11, no. 2, pp. 37–50, 1912.

[30] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE