Comparison of Different Convolutional Neural Networks Utilizing Transfer Learning for Pneumothorax Segmentation from Whole Chest X-Ray Images and Extracted Patches

Lazar Dašić[1,2 [0000-0002-8055-100X]], Ognjen Pavić[1,2[0000-0003-2533-1079]], Tijana Geroski[2,3[0000-0003-1417-0521]], Mina Vasković Jovanović[3[0000-0003-3644-8723]] and Nenad Filipović[2,3[0000-0001-9964-5615]]

[1] Institute for Information Technologies Kragujevac, University of Kragujevac, Jovana Cvijića, 34000 Kragujevac, Serbia
[2] Bioengineering Research and Development Center (BioIRC), Prvoslava Stojanovića 6,34000 Kragujevac, Serbia
[3] Faculty of Engineering, University of Kragujevac, Sestre Janjića 6, 34000 Kragujevac, Serbia
lazar.dasic@kg.ac.rs

Abstract. Pneumothorax is a lung condition characterized by the presence of air between chest wall and lungs. In order to diagnose location and size of pneumothorax, chest X-ray is a commonly used imaging technique. U-Net convolutional neural network models with different backbones are compared in order to assess their capability to automatically and correctly segment signs of pneumothorax from chest X-rays. Five different pretrained backbones have been chosen: VGG19, ResNet34, ResNet50, DenseNet121 and Inceptionv3. Two different approaches for pneumothorax segmentation have also been tested: one methodology used X-ray images of the whole chest area for training, while the second one split the original images into patches and used them for the training process. Both methodologies performed at a similar level, with the best results achieved by U-Net model with DenseNet121 backbone for segmentation of X-ray of the whole chest. This model achieved a Jaccard index and Dice score of 76.92% and 78.81%, respectively. These results indicate that the tested models are capable of extracting fine-grained features from X-ray images of whole chest and that patch-based segmentation does not provide additional benefits.

Keywords: chest X-ray, convolutional neural networks, pneumothorax segmentation.

1 Introduction

Pneumothorax is a lung condition characterized by the presence of air in the pleural cavity (area between chest wall and lungs) [1]. The existence of air in this area interferes with the negative pressure that normally prevents the lungs from collapsing, which can lead to a partial or complete collapse of the lung that can potentially be fatal. Depending on their cause, pneomothoraces can be classified as [2]:

- Primary spontaneous pneumothorax – This pneumothorax occurs suddenly in healthy individuals with no traces of underlying lung disease present.
- Secondary spontaneous pneumothorax – This pneumothorax occurs due to the present underlying lung disease, such as chronic obstructive pulmonary disease (COPD), lung infections (e.g. tuberculosis), interstitial lung disease, etc.
- Traumatic pneumothorax – This pneumothorax occurs as a result of physical trauma to the chest and lungs. Traumas can come from various sources, ranging from blunt force traumas (e.g., car crashes) to penetration injuries (e.g., stab wounds or gunshot wounds).
- Iatrogenic pneumothorax – This pneumothorax occurs as a direct result of certain medical procedures, such as central vein cannulation, needle biopsy, pleural tap, etc.

Although the exact number of patients affected by pneumothorax is unknown, the incidence of primary spontaneous pneumothorax is reported to be 18-28/100,000 cases per year for men and 1.2-6/100,000 for women [3]. The dangers and subsequent treatments of pneumothorax are directly correlated to its size. The Thoracic Society recommends classifying the size of a pneumothorax based on the visible space between the edge of the lung and the chest wall, thus having two categories [4]:

- small pneumothorax (less than 2 cm present between lungs and chest wall)
- large pneumothorax (more than 2 cm present between lungs and chest wall)

Small pneumothorax areas are usually healed without treatment, as the excess air will be absorbed over time. As the size of pneumothorax area increases, the treatment methods become more invasive, where using procedures such as needle aspiration, chest drain and even surgery becomes common.

Even though it is possible to diagnose pneumothorax using only physical examination and patient's medical history, medical imaging is almost always used for a definitive diagnosis and determination of size and severity of a pneumothorax. Due to its relatively low cost, ease of use and wide availability, chest X-ray is primarily used medical imaging modality. On chest X-rays, there are a couple of signs that may indicate presence of pneumothorax: visible pleural line, absence of lung markings peripheral to pleural line and presence of radiolucent area. However, in the case of a small pneumothorax these sings may be hard to detect, resulting in doctor's failure to notice presence of pneumothorax.

In order to reduce a workload and assist doctors in interpretation of medical data, many Computer-Aided Diagnosis (CAD) systems have been developed. In the case of pneumothorax diagnosis, having a system that automatically and precisely finds visual sings of pneumothorax on X-rays, could significantly lessen the burden on medical experts and assist them in clinical practice. In the assessment of pneumothorax through medical imaging, it is not enough to simply identify the existence of air within the pleural space. Medical experts need to know the pneumothorax's position and size to evaluate the severity of the condition and provide appropriate treatment for the patient. The task of annotating pneumothorax in X-ray images is categorized as a binary segmentation problem, with the primary goal being the detection and delineation of the region-of-interest (ROI).

Development and use of Convolutional Neural Networks (CNNs), a subset of deep learning techniques, have led to considerable progress in the field of medical imaging, especially for the task of image segmentation. Deep learning methods are capable of automatically extracting complex features at various scales without manual intervention. For the task of image segmentation, CNN models with U-Net architecture achieve excellent results and are often use in analysis of biomedical data.

The goal of this research is to compare U-Net models with different pretrained backbones and assess their performance for the task of pneumothorax segmentation from chest X-ray images. Two different methodologies are used to train models: one uses X-rays of the whole chest area for training, while the second splits the X-ray into small patches for model training. The idea of using patches for pneumothorax segmentation was inspired by the results achieved in similar tasks in medical field [5, 6]. We argue that performing segmentation using smaller patches would result in better segmentation results, especially for detection of small pneumothoraces that can be difficult to detect due to their size and variation in positioning. The results of the patch segmentation would then be stitched together to reconstruct a segmentation mask that displays the entire chest area.

1.1 Related works

Segmentation of pneumothorax from chest X-ray images using deep learning approach has been a popular research topic, with numerous proposed solutions emerging. Most of these deep learning solutions use CNNs, especially the pretrained kind.

Malhtora et al. [7] proposed Mask Regional Convolutional Network (Mask RCNN) with pretrained ResNet101 backbone. Their model generates bounding boxes as well as segmentation masks for every instance of an object present in the given image.

Wang et al. [8] created methodology consisting of two stages: firstly, presence of pneumothorax is detected in the image and in second phase, ensemble of four U-Net like models (with pretrained SEResnext50, SE-Resnext101, EfficientNet-B3, and EfficientNetB5 backbones) and one Deeplabv3+ model is used for segmentation. By doing classification of X-rays into healthy and pneumothorax categories, authors avoided any errors that could occur if a segmentation model incorrectly identified pneumothorax in a healthy patient's X-ray.

Usage of ensemble methods seem to be popular idea, as Abedalla et al. [9] use ensemble of pretrained ResNet50, DenseNet169, SE-ResNeXt50 and EfficientNet-B4 U-Net models.

2 Materials and Methods

In this section, methodology for development and comparison of different CNN models is described. Five different pretrained models have been trained on both X-rays of the whole chest area and extracted patches, resulting in total of ten different models. Figure 1 shows the proposed methodology workflow.

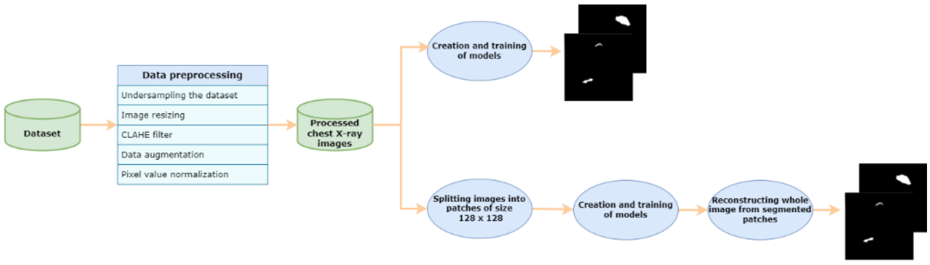

Fig. 1. Proposed methodology workflow

2.1 Dataset description

Chest X-ray images that were used for development and evaluation of segmentation models are a part of public SIIM-ACR pneumothorax dataset. This dataset was used in a Kaggle challenge hosted by The Society for Imaging Informatics in Medicine (SIIM) and the American College of Radiology (ACR) [10]. The dataset consists of 12,047 chest X-rays with corresponding annotated masks, among which 2,669 images have pneumothorax present, and the other 9,378 images are of healthy lungs. Figure 2 shows some examples of X-ray images with corresponding annotated masks, where white sections represent pneumothorax instances and totally black masks represent healthy individuals.

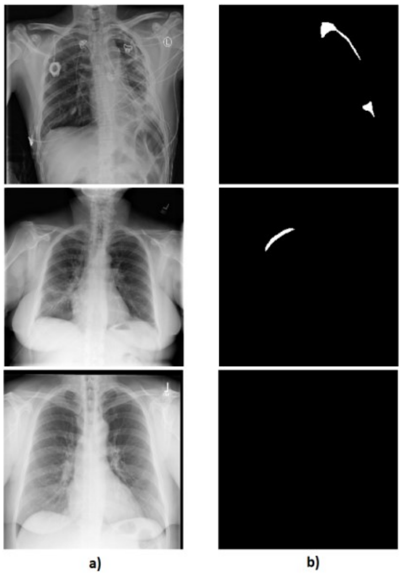

Fig. 2. Examples of dataset's image pairs: a) Original chest X-ray image; b) Annotated pneumothorax

2.2 Data preprocessing and augmentation

Looking at the SIIM-ACR dataset's metadata it is evident that samples are highly imbalanced, with only 22.15% of images belonging to the pneumothorax class. Having class imbalance in the training dataset is significant problem that often causes biased training and poor generalization. In order to address this issue, undersampling of the majority class was utilized: all of the samples containing pneumothorax were used, while 2,669 X-rays of healthy patients were randomly selected.

Original images have dimensions of 1024 × 1024 pixels. Due to the memory restrictions and computational efficiency, images have been halved to a size of 512 × 512 pixels. In order to improve contrast of the image, which would make radiolucent area of pneumothorax more prominent, Contrast Limited Adaptive Histogram Equalization (CLAHE) filter was used.

In order to improve diversity of training data, combination of following image transformations has been used to artificially expand the dataset: random rotation by 0-20°, random change of brightens by ±0-15%, random zoom by ±0-10% and horizontal flip.

Final step in data preprocessing is normalization of pixel values for both images and masks. Usual pixel values range from 0 to 255 and normalization brings all images to a consistent scale between 0 and 1.

The preparation of the dataset for image segmentation of whole chest area encompasses all the outlined steps. Subsequently, this dataset was then split into two subsets, with 80% of samples were used as training dataset and the remaining 20% serve as the test dataset. However, for patch-based methodology original images had to be split into smaller patches. Patch size of 128 × 128 pixels was selected, as this size allows capturing of fine-grained features while still being able to capture largest objects without losing too much context.

2.3 Methods

Convolutional neural networks are considered state-of-the-art for tasks in computer vision, such as image classification and segmentation, object detection, etc. A typical CNN architecture consists of sequential blocks composed of different layers, with each block designed to transform an input volume into an increasingly abstracted output. At the core of CNN architecture is a convolutional layer which is capable of extracting hidden features of different scale, where early convolutional layers in early blocks learn low-level features (e.g. colors, brightness, etc.) and deeper ones learn complex patterns. A significant advantage of CNNs over traditional fully connected networks is their efficiency in processing visual data. This efficiency is partially achieved through the use of pooling layers, which reduce the spatial dimensions of feature maps. By doing so, pooling layers decrease the number of parameters and computational load of network, which not only enhances efficiency but also helps mitigate overfitting.

All of the models created in this research follow structure of a U-Net convolutional neural network. U-Net represent a huge milestone in development of CNNs for segmentation and it still remains one of the most commonly used deep learning architectures for a variety of segmentation tasks. U-Net is an architecture developed specifically for biomedical image segmentation by Ronneberger and his team [11]. U-Net comprises of two parts, encoder and decoder, which form U-shaped layout that the architecture is named after. The encoder, or contracting path, is standard CNN consisting of blocks of sequential convolutions and max pooling layers. Encoder is tasked with extraction of relevant features from images while reducing size of input images with pooling operations. During the feature extraction, original size of images and position of pixels is lost, which is why there is a need for a decoder. Decoder, or expanding path, using up-convolutions and skip connections combines the encoder's extracted features with spatial information to create a segmentation mask. Skip connections are crucial to the U-Net architecture as they propagate high-resolution information from earlier encoder layers to later blocks in decoder.

Inspired by popularity and success of U-Net, numerous modified versions of U-Net structure that improve its performance have been proposed. In order to improve feature extraction capabilities of U-Net encoder, state-of-the-art backbones are used as a replacement for standard CNN consisting of blocks of sequential convolutions and max pooling layers. These improved backbones, such as ResNet, DenseNet, etc., can make U-Net architecture more suitable for extraction of features specific to certain problems. For the task of pneumothorax segmentation, VGG19, ResNet34, ResNet50, DenseNet121 and Inceptionv3 backbones have been tested for both segmentation of X-rays of whole chest area and segmentation of patch images.

VGG architecture was developed by Simonyan et al. [12], representing an important step in development of CNN models, advocating for increase in networks' depth and replacement of large convolutional filters with many smaller 3×3 filters. These ideas have become staple and have been used in development of many future models. VGG has several variations, most notably VGG16 and VGG19 that contain 16 and 19 layers, respectively.

He et al. [13] created Residual Network (ResNet) that introduces residual blocks to CNN structures. Residual blocks utilize skip connection to bypass one or more layers, resolving vanishing gradient by facilitating the flow of gradients during backpropagation. Consequently, they enable the successful training of much deeper neural network models, such as those with 34, 50, 101, and even 152 layers, without performance degradation often encountered in conventional deep CNNs.

Traditional CNNs pass the output of one layer directly into the next layer, but developers of DenseNet had different approach and created a model whose every layer is connected to every other subsequent layer [14]. This densely connected architecture allows the network to reuse features throughout the depth of the network, which can lead to improved efficiency and feature propagation.

Most of the aforementioned models aimed to increase performance by going deeper, adding more layers to their networks. However, the Inceptionv3 architecture introduced the novel concept of also expanding the network width [15]. Instead of using convolutional layers with only filters of the same size, the Inceptionv3 architecture incorporates Inception modules that apply several convolutional filters with different sizes (1×1, 3×3, 5×5, etc.) in parallel. This allows capture of features of different scale simultaneously.

All of the selected U-Net backbones have been pretrained on the ImageNet dataset. The goal of the transfer learning procedure is to utilize knowledge gained while solving one problem and apply it to a different problem that is related. The pretrained model provides initial values for the training parameters and obviates the need for training from scratch.

In order to compare selected models as fairly as possible, same training hyperparameters have been used for all of them. The models were trained for 75 epochs with a batch size of 8 using Adam optimizer with a learning rate of 0.001. To mitigate the effects of the imbalanced dataset, which has a large number of background pixels, a custom weighted loss function that combines categorical focal loss with dice loss was utilized.

The proposed models were developed using Python programming language and Tensorflow library. Model training was conducted on a workstation equipped with a NVIDIA GeForce GTX 1660 GPU, an AMD Ryzen 7 2700X CPU and 16 GB of RAM.

3 Results and discussion

Evaluation of segmentation models involves use of various quantitative metrics that measure the quality of model's prediction. For the task of image segmentation most commonly used metrics are Intersection-over-Union (IoU) and Dice similarity coefficient (DSC). IoU, or Jaccard Index, measures ratio between the intersection and the union of the predicted and ground-truths segments. IoU is calculated as:

$$\mathrm{IoU} = \frac{|X \cap Y|}{|X \cup Y|} \quad (1)$$

, where X represent resulting segmentation mask and Y is annotated ground-truth. DSC is similar to IoU, as it also measures the overlap between two samples. DSC is calculated as:

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

The results of segmentation on the test set, for all of the created models are shown in Table 1.

Table 1. Evaluation of trained models using IoU and DSC metrics

U-Net backbones	Segmentation of X-rays of the whole chest		Patch-based segmentation	
	IoU [%]	DSC [%]	IoU [%]	DSC [%]
VGG16	70.33	72.41	71.02	72.96
ResNet34	73.54	75.36	72.23	73.47

ResNet50	74.28	76.21	72.45	74.12
DenseNet121	76.92	78.81	75.73	77.24
Inceptionv3	74.92	77.12	75.26	77.89

Evaluation results from Table 1 show that models perform at similar level, with DenseNet121 performing best for both segmentation of X-rays of whole chest area and segmentation of patches. Results also indicate that using smaller patches of chest X-rays does not provide significant improvement in pneumothorax segmentation. Figure 3 shows the results achieved by U-Net model with DenseNet121 backbone for both whole image and patch-based segmentation.

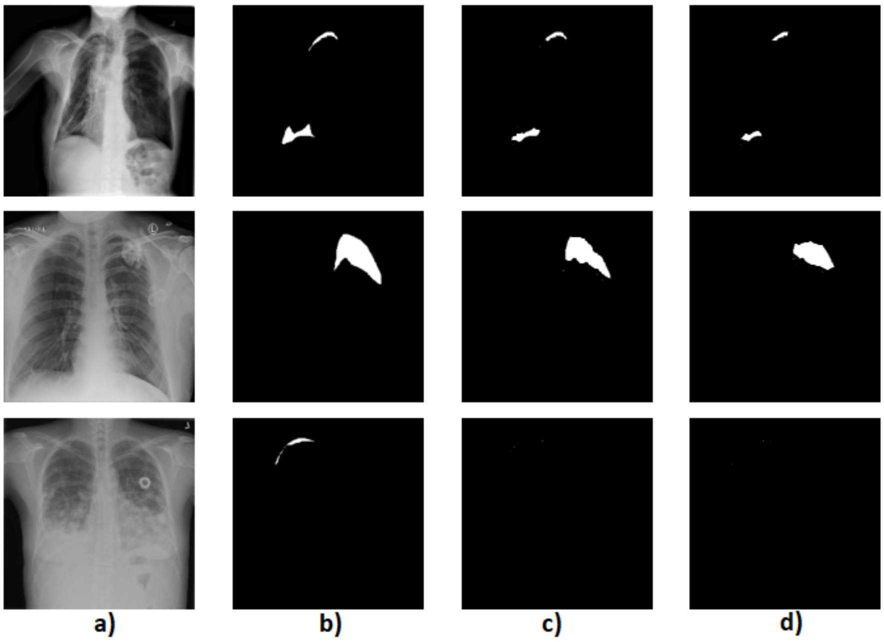

Fig. 3. Segmentation result of a U-Net model with DenseNet121 backbone: a) original image, b) annotated ground-truth; c) whole image segmentation results; d) patch-based segmentation results

4 Conclusions

This paper compared performance of U-Net models with different backbones for the task of pneumothorax segmentation. Five different backbones have been tested: VGG19, ResNet34, ResNet50, DenseNet121 and Inceptionv3. The task of pneumothorax segmentation was approached in two different manners: one methodology used X-ray images of the whole chest area for training, and the second one split the original images into patches and used them for the training process.

DenseNet121 backbone had best performance for both approaches, as evident by IoU and DSC scores. However, results comparing the two methodologies seem to be similar for all of the tested models. This indicates that splitting images into smaller patches, segmenting them, and then stitching them back together to reconstruct the mask of the whole chest area, does not provide any benefits over doing segmentation on the X-ray of the whole chest area.

In future research, chest X-ray images are going to be subjected to additional step in data preprocessing. Idea is to use additional CNN that is going to segment lungs from the surrounding tissue and background. These images that contain just the lung area would be used as input for training an DenseNet121-based U-Net model for pneumothorax segmentation, thus putting focus only on that area during training procedure.

5 Conflict of interest

Authors have no conflict of interest to declare.

6 Acknowledgements

This research is funded by the project that has received funding from the Euro-pean Union's Horizon 2020 research and innovation programme under grant agreement No 101016834 (HOSMARTAI). This research was funded by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, contract numbers [451-03-66/2024-03/200378 (Institute for Information Technologies, University of Kragujevac) and 451-03-47/2023-01/200107 (Faculty of Engineering, University of Kragujevac)].

References

1. M. Noppen and T. De Keukeleire, "Pneumothorax," *Respiration,* vol. 76, no. 2, pp. 121-127, 2008.
2. G. P. Currie, R. Alluri, G. L. Christie and J. S. Legge, "Pneumothorax: an update," *Postgraduate medical journal,* vol. 83, no. 981, pp. 461-465, 2007.
3. L. J. Melton III, N. G. Hepper and K. P. Offord, "Incidence of spontaneous pneumothorax in Olmsted County, Minnesota: 1950 to 1974," *American Review of Respiratory Disease,* vol. 120, no. 6, pp. 1379-1382, 1979.
4. M. Henry, T. Arnold and J. Harvey, "BTS guidelines for the management of spontaneous pneumothorax," *Thorax,* vol. 86, no. Suppl 2, p. ii39, 2003.
5. R. Mechrez, J. Goldberger and H. Greenspan, "Patch-based segmentation with spatial consistency: application to MS lesions in brain MRI," *Journal of Biomedical Imaging,* vol. 2016, p. 13 pages, 2016.

6. S. Raghavendra, E. B. Dam, S. A. Flensborg and J. Petersen, "Patch-based Medical Image Segmentation using Matrix Product State Tensor Networks," *arXiv preprint,* vol. arXiv:2109.07138, 2021.
7. P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, M. Kaur and H. N. Lee, "Deep learning-based computer-aided pneumothorax detection using chest X-ray images," *Sensors,* vol. 22, no. 6, p. 2278, 2022.
8. X. Wang, S. Yang, J. Lan, Y. Fang, J. He, M. Wang, J. Zhang and X. Han, "Automatic segmentation of pneumothorax in chest radiographs based on a two-stage deep learning method," *IEEE Transactions on Cognitive and Developmental Systems,* vol. 14, no. 1, pp. 205-218, 2020.
9. A. Abedall, A. Abdullah, M. Al-Ayyoub and E. Benkhelifa, "Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures," *PeerJ Computer Science,* vol. 7, p. e607, 2021.
10. A. Zawacki, C. Wu, G. Shih, J. Elliott, M. Fomitchev, M. Hussain, P. Lakhani, P. Culliton and S. Bao, "SIIM-ACR Pneumothorax Segmentation," Society for Imaging Informatics in Medicine, [Online]. Available: https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation. [Accessed 25 01 2024].
11. O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention,* pp. 234-241, 2015.
12. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint,* vol. arXiv:1409.1556, 2014.
13. 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.
14. G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
15. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.