Cycle GANs for Lazy Label Segmentation of Medical Imaging

Cale Gregory
MIT
caleg@mit.edu

Abstract

I explore segmentation of the lesion tissue and afflicted areas in medical imaging with image-wise labeling using domain transfer. Usually training for segmentation utilizes datasets that have pixel level labels of afflicted and healthy tissues. Unfortunately, these datasets are difficult to find and onerous to create; therefore, using picture level labels and unsupervised learning of the data would be ideal. Since healthy medical imaging contains all components of unhealthy imaging except for the afflicted area, one can use a domain transfer algorithm to remove the afflicted tissue from the image. Then, the unhealthy tissue can be segmented by subtracting the fake healthy image from the real sick image. To implement this, Cycle GANs (Generative Adversarial Networks) allow for more freedom as the healthy images do not have to be from previously scanned sick patients.

1. Introduction

Chest radiographs are the most common medical imaging in the world, yet there are not enough radiologists to serve all those that need these images [1]. To automate the process of reading chest x-rays, segmentation can be used to show where afflicted tissue is on the lungs or other organs. Convolutional neural networks (CNNs) have been hugely successful in analyzing and classifying pictures in many fields including medical [1]. However, common classification models such as Res-Net tend to have blob-like and inaccurate segmentations of the afflicted areas [2]. There have been multiple attempts to make this segmentation more precise by adding smoothing factors through conditional random fields as recurrent neural networks (CRF-RNN) or utilizing fully convolutional nets such as U-Nets [2, 3]. U-Nets have pixel-wise segmentation accuracy of up to ninety-two percent; however, these models need to be trained on pixel-wise labeled datasets.

To circumvent the need for heavily labeled datasets, semi-supervised learning can be used. Networks can differentiate between classes of pictures based on the features within. Therefore, prelabeled imaging as either diseased or healthy is all the computer needs to learn difference between these domains. For example, a healthy lung will not have fluid inside, where as an unhealthy one might. Utilizing unsupervised learning and domain transfer networks, the fluid should be removed from the lung during the domain transfer. These differences between the diseased and healthy domains should segment the afflicted areas within the image.

2. Related Work

GANs are commonly used in the creation of new images to supplement datasets that lack a substantial number of cases for certain diseases [4, 5]. For example, cycle GANs have been used to study the binary classification between a chest x-ray that is healthy or suffering from pneumonia [5]. In addition, these processes can be performed with minimal expert knowledge as image-wise labeling does not require in depth knowledge of how to read medical imaging. This information can be gleaned from the medical reports that tend to be attached to these datasets.

Other forms of data augmentation have been documented to be successful for medical imaging. In a case study on COVID-19, neural networks were better at classifying images if trained on datasets that randomized flips along the y-axis, translations, and rotations [6]. Since I am performing a case study on chest radiology, the same data augmentation methods can be used to further strengthen my dataset.

Cycle GANs have been used for semantic segmentation and have shown around a sixty-two percent (62%) pixel to pixel accuracy [7]. In the study, the domains consisted of the segmented domain and the non-segmented domain, so the creation of the pixel-wise dataset is needed to train these generators [7]. However, the use of healthy and sick domains provides the opportunity to utilize the high similarities between the healthy and diseased domain. Therefore, the high similarity in the features that are insignificant in the segmentation allow for a similar process to be used for segmentation.

3. Method

3.1 Cycle GAN Structure

Cycle GANs consist of two GANs (a generator network and a discriminator network) and work as an image domain transfer mechanism. For medical imaging, the domains of healthy imaging and unhealthy imaging are used. A cycle GAN is trained for each disease that is segmented. Ideally, one can diagnose the disease prior to segmentation to choose the proper domain transfer generator. On similar datasets, this classification task can be completed with ninety-three percent accuracy by today's models [1].

Each cycle GAN utilized generators that expand the feature space by a factor of two for three convolutional layers with a kernel size of 3x3 and ReLU activation and batch normalization with kernel size of 2x2. In addition, the generators had six residual blocks to add non-linearities to the mutation of the original image. The final image is then decoded by halving the feature space for three transposed convolutional layers to return to the original feature space of the image. The discriminators consisted of four convolutional layers with batch normalization and leaky ReLU activation.

3.2 Thresholding

To segment the afflicted tissue, one can use the sick domain to healthy domain generator to mark the differences between the two images. To get the difference matrix, the new and fake healthy image is subtracted from the original unhealthy image. For final segmentation, scale the difference matrix so that its values exist between 0 and 1. Then, threshold the outliers away. I found that the best results come from the threshold of greater than 0.3 and less than 0.8.

The formula for finding the differences in the domains can be written as

$$|X - X'| / \max(X - X')$$

such that X is the original image and X' is the image after domain transfer.

4. Results

4.1 Datasets

I utilized two datasets that contained picture wise labels depending on the disease or lack there of in the chest x-ray. Both datasets are available on Kaggle and named "Chest X-Ray Images (Pneumonia)" and "Random Sample of NIH Chest X-ray Dataset". The former contains a split dataset of 5856 chest x-rays with a fifty-fifty split between pneumonia-afflicted and healthy radiographs. The latter contained a set of 14 diseases. I utilized the data for hernia, emphysema, and edema to diversify the types of diseases during the experiments.

4.2 Feature Space and Pneumonia

I began by experimenting on pneumonia to optimize the model faster than testing it on lots of diseases. I trained two models depending on size of the original images. The first was a height of 64 and resizing of width while maintaining the original aspect ratio of the scan. The model trained for 500 epochs on the pneumonia dataset from Kaggle. The generators and discriminators used the Adam optimization with a learning rate 0.005 and 0.0002 respectively. I augmented the data with a random rotation of up to 5 degrees and a random flip along the horizontal axis of the training data. This model failed to segment the inflammation caused by pneumonia in a meaningful way as many of the features were lost through the down-sampling as pictured by Figure 1.

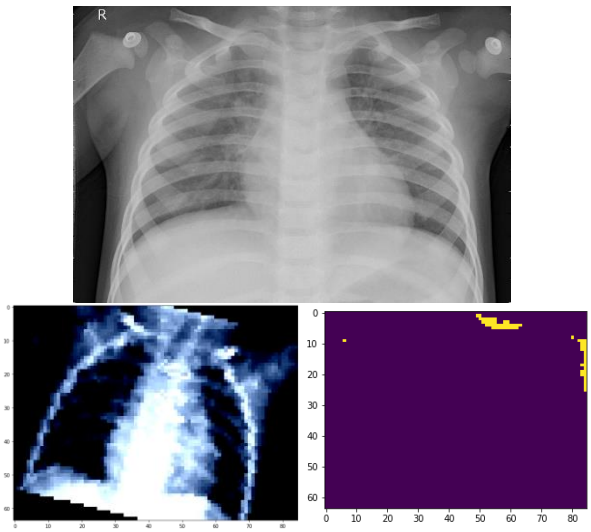

Figure 1: Example of complexity loss during down sampling and failed segmentation. Original image (top), down sampled image (bottom left), and attempted segmentation (bottom right).

To remedy this loss of complexity, I extended the model to take images of 128 pixels high. This allows for smaller features to be present in the encoding. As seen in Figure 2, the cycle GAN recognized the features that disappeared in the other model.

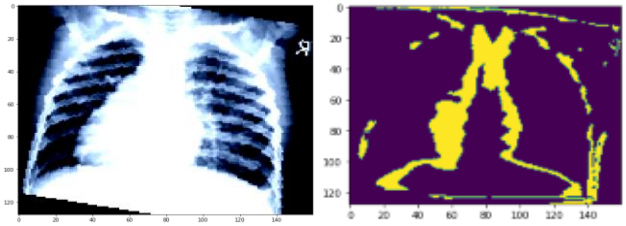

Figure2: Down sampled radiograph of a pneumonia -afflicted chest (left) and segmentation through proposed method (right).

While some complexity is still lost in the preprocessing for the segmentation, it is the largest dimensional input that I was able to train on the GPU as the model would grow to large if the initial image dimension were increased.

4.3 Expanding Diseases

To test if the difference between the domains truly worked to segment afflicted tissue, I expanded my set of experiments to other medical traumas such as edema, emphysema, and hernias. To test the accuracy of my system, I created a small validation set for each disease and constructed a ground truth image for each image in the validation set. From there, I found the pixel-to-pixel accuracy of the segmentation by taking the mean squared error of the ground truth and the proposed segmentation.

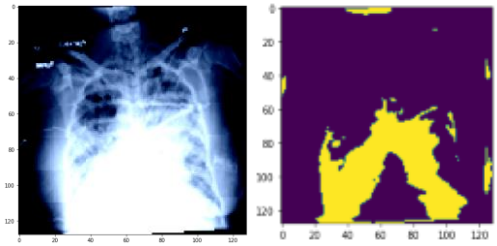

Figure 3: chest radiograph of edema patient (left) and segmentation of edema (right)

I trained a cycle GAN for each category. Every generator and discriminator were optimized using Adam with a learning rate of 0.005 and 0.0002 respectively for 500 epochs. As seen in Figure 4, pneumonia and edema were segmented the best out of the diseases.

Disease	Pneumonia	Hernia	Emphysema	Edema
64 px	49.77			
128 px	57.47	47.96	51.75	57.38

Figure 4: Pixel-to-pixel accuracy of the segmentation to ground-truth

4.4 Failure Cases and Limitations

Hernias were a place of failure for the algorithm. A hernia causes a shift in the placement of the features (organs, muscles, and spinal cord) rather than presenting a foreign body or inflaming an already existing one to the radiograph. As a result, the domains were too close to make a noticeable difference. Therefore, images would like almost identical on each side of a generator and ruining the segmentation technique as seen in Figure 5.

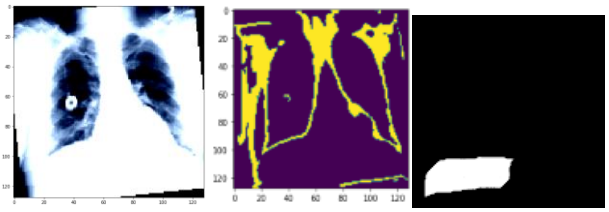

Figure 5: Original radiograph of hernia (left), attempted segmentation (middle), ground truth of displaced tissue (right)

In addition, the domain transfer only segments the difference between a healthy and unhealthy image; therefore, it does not differentiate the disease that is being segmented. This can be remedied by having a domain strictly without a specific disease and with a specific disease. Then, the same method can be used on the image by several generators from these cycle GANs trained on different diseases to segment different classes.

5. Conclusion

In this project, I investigated the ability to segment chest radiographs without the need for pixel-wise labeling. The goal was to utilize unique characteristics of medical imaging to allowing for lazy labeling of datasets for segmentation. As a result, cycle GANs for segmenting the difference between the healthy and diseased domains of medical images provide a more accessible method of segmentation than creating pixel-wise labels. Using the case study of chest radiographs, the method provided astonishing results of segmenting pneumonia and edema. However, it fails at recognizing relative positioning of features; therefore, it is limited to diseases that cause inflammation, abnormal growths, or foreign bodies. Future studies can be conducted on other forms of medical imaging as the segmentation of diseases with minimal labeling had promising results on chest radiographs.

References

[1] J. Irvin and P. Rajpurkar and M. Ko and Y. Yu and S Ciurea-Ilcus and C. Chute and H. Marklund and B. Haghgoo and R. Ball and K. Shpanskaya and J. Seekins and D. A. Mong and S. S. Halabi and J. K. Sandber and R. Jones and D. B. Larson and C. P. Langlots and B. N. Patel and M. P. Lungren and A. Y. Ng, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison," *arXiv preprint arXiv:1901.07031*, 2019.

[2] S. Zheng and S. Jayasumana and B. Romera-Paredes and V. Vineet and Z. Su and D. Du and C. Huang and P. H. S.s Torr., "Conditional Random Fields as Recurrent Neural Networks," *arXiv preprint arXiv:1502.03240*, 2015.

[3] O. Ronneberger and P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv preprint arXiv:1505.04597*, 2015.

[4] Maayan Frid-Adar et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. CoRR, abs/1803.01229, 2018.

[5] T. Malygina and E. Ericheva and I. Drokin, "GANs 'N Lungs: improving pneumonia prediction," *arXiv preprint arXiv:1908.00433*, 2019.

[6] M. Elgendi and M. U. Nasir and Q. Tang and D. Smith and J. Grenier and C. Batte and B. Spieler and W. D. Leslie and C. Menon and R. R. Fletcher and N. Howard and R. Ward and W. Parker and S. Nicolaou, "The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective," *Frontiers of Medicine*, vol. 8, 2021.

[7] C. Zhang and Y. Tang and C. Zhao and Q. Sun and Z. Ye and J. Kurths, "Multi-task GANs for Semantic Segmentation and Depth Completion with Cycle Consistency," *arXiv preprint arXiv:2011.14272*, 2020.