CS482/682 Final Project Report Group 20

Detection and Segmentation of Pneumothorax from Chest X-ray Images

Ameya Harmalkar, Yixuan Huang, Simon Liu, Fangchi Shao

1 Introduction

Background Pneumothorax (PTX), colloquially known as a collapsed lung, is a condition where air leaks into the space between the lung and chest wall and can be challenging to diagnose from chest X-ray (CXR) images [1]. We intend to develop a deep learning algorithm that enables automatic PTX segmentation to provide reliable diagnoses. However, densely labeled CXR images require extra effort from experts and are not typically generated in clinical settings. Therefore, transfer learning from sparsely labeled images can be an effective method to increase a network's CXR exposure. Furthermore, unlabeled images can be included in an unsupervised manner, which enables more images to contribute to the final segmentation task. In this work, we study the effect of transfer learning from hospital-scale CXR images.

Related Work U-Nets have shown promising performance in medical image segmentation by avoiding vanishing gradients and protecting upsampling information. VGG, ResNet and other classfiers have been used as U-Net encoders. For transforming images into low-dimensional latent space, autoencoders (AE) and Variational-AEs [2] can serve as potential tools.

2 Methods

Dataset The dataset is composed of X-ray DICOM images and annotations in the form of image IDs and run-length-encoded (RLE) masks. PTX-positive images are indicated by encoded binary masks while PTX-negative images have a mask value of -1. We trained our model for PTX prediction in test images and indicated PTX locations with binary masks. Since official test data is not public, a training subset was used as the test set (N=1159).

Segmentation A U-Net with a ResNet-34 encoder [3] is used to perform the segmentation task. The loss function is designed as a weighted sum of focal loss and dice loss [4]. Focal loss is included to mitigate the imbalance between background and foreground objects, as the background contributes to the loss much more than PTX. The dice score is computed using the criteria given in the project evaluation. The network is trained with Adam and learning rate of 1e-4. The images are downsampled to 512x512 pixels and trained with batch size of 8. To improve generalizability, augmentations like random flipping and affine transforms are applied.

Supervised Transfer Learning To improve segmentation performance, we investigated the effects of transfer learning from weakly labeled CXR datasets. A hospital-scale chest X-ray database [5] with 14 different disease labels was used for pre-training the U-Net encoder. A ResNet-34 model was trained to classify if the image contains PTX, which then replaced the U-Net encoder and trained with the SIIM dataset for segmentation. Since PTX-positive class proportion were small, a weighted sampler was used

Figure 1: CXR image (1) reconstructed by the Autoencoder (2) and Variational Autoencoder (3).

to oversample the positive samples and mitigate class imbalance. Due to time constraints, the training was conducted on 10000 images and evaluated on 1000 images with cross-entropy loss, SGD optimizer, learning rate of 1e-3 and momentum of 0.9. The learning rate was adjusted per validation loss, which reduces by a factor of 10 if the validation loss does not decrease for 5 epochs. The network was trained until convergence and then used as the U-Net encoder.

Unsupervised learning In order to emulate unlabeled data, we disregarded the weakly labeled CXR labels and performed unsupervised clustering to generate pseudo-labels. This was done by training an AE (with BCE Loss) and a VAE (β-VAE with ELBO loss) on the unlabelled CXRs to extract a set of deep features in the encoded latent space. Following training, we use Density-based Spatial Clustering of Applications with Noise (DBSCAN) for constructing 14 pseudo-clusters. The network was trained for over 50 epochs until loss convergence. The reconstructed images are illustrated in Figure 1. We used PCA on the encoding to reduce the dimensions and cluster into 14 distinct pseudo-labels that is fed into the ResNet34 encoder for pre-training.

3 Results

Pre-training Classification With true labels, the network reached an accuracy of 77% on 2-class classification (PTX vs. no PTX). With pseudo-labels, the network reached 31.91% accuracy on the 14-class classification.

Segmentation The U-Net segmentation results with different encoders are shown in Table 1.

Model	Dice score
No additional pretrain	71.6
Pretrain with labels	76.4
Pretrain with psuedo-labels	75.8

Table 1: Segmentation performance of models with different pre-training.

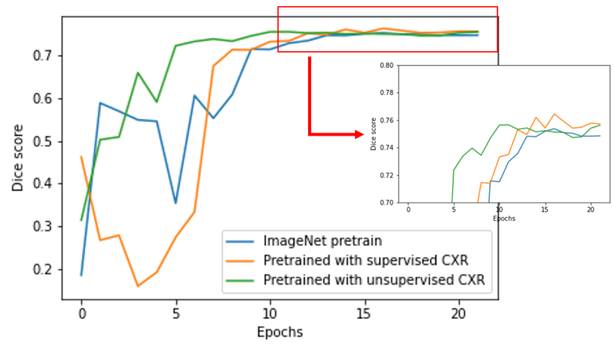

Figure 2: Dice score curve of different pre-trainings.

4 Discussion & Conclusion

We demonstrated that pre-training improved the segmentation performance by 5% (Table 1 & Figure 2). Due to computational resource constraints, we did not use the full weakly labeled dataset. Future work could focus on direct incorporation of the AE into the U-Net without ResNet-34 pretraining.

References

[1] I Satia et al. "Assessing the accuracy and certainty in interpreting chest X-rays in the medical division". In: *Clinical medicine* 13.4 (2013), p. 349.

[2] Asmaa Abbas, Mohammed M. Abdelsamea, and Mohamed Medhat Gaber. "4s-dt: self supervised super sample decomposition for transfer learning with application to Covid-19 detection". In: *arXiv* 14 (2020). DOI: 10.1101/2020.06.22.20137547. arXiv: 2007.11450.

[3] Pavel Yakubovskiy. *Segmentation Models Pytorch*. https://github.com/qubvel/segmentation_models.pytorch. 2020.

[4] Rishabh Agrahari. *UNet with ResNet34 encoder*. https://www.kaggle.com/rishabhiitbhu/unet-with-resnet34-encoder-pytorch. 2019.

[5] Xiaosong Wang et al. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2097–2106.