Final report

Siddharth Ranjan
Arizona State University
Tempe, Arizona 85288
sranja18@asu.edu

Abstract

This report discusses the integration of classification, localization and segmentation using Swin transformers and Faster CNN on ChestX-ray14 and TBX11K datasets.

1. Introduction

1.1. Swin Transformer

Swin Transformer is a vision transformer which is used for image classification, localization. It uses a hierarchical structure and shifted window mechanism. Shifted window mechanism helps to capture relationships between different parts of the image. Hierarchical structure is useful for downsampling.

1.2. ChestX-Ray 14

ChestX-ray 14 is large dataset which contain images of chest x-rays. It has 112000 images. It has 14 disease labels of thoracic diseases like pneumonia, hernia etc. It is widely used to train convolutional neural networks and transformers.

1.3. TBX11-K

TBX11-K consist of chestx-ray images for tuberculosis. The dataset also consists of annotations for localization.

1.4. Faster R-CNN

Faster R-CNN is a deep learning model designed for efficient and accurate object detection. It integrates a Region Proposal Network (RPN) to rapidly generate possible regions containing objects, followed by a classifier and bounding box regressor to accurately identify and localize objects within these regions. This approach allows for quicker processing compared to older models like R-CNN and Fast R-CNN, without sacrificing detection precision. While it may not achieve real-time speeds like YOLO, Faster R-CNN offers a strong balance of accuracy and speed, making it suitable for applications where precise object detection is crucial, even if a slight delay is acceptable.

1.5. ChestX-Det

The ChestX-Det dataset is a carefully curated set of annotated chest X-ray images aimed at supporting the development and evaluation of deep learning models for thoracic disease detection. It provides a comprehensive collection of chest radiographs labeled with various abnormalities, such as pneumonia, lung opacities, and enlarged heart conditions, allowing for the training of models that can identify and localize these issues. The dataset is valuable for advancing research in automated diagnostic tools, enabling machine learning models to assist radiologists by highlighting areas of concern and improving diagnostic accuracy in clinical settings.

1.6. U-Net

U-Net is a deep learning architecture widely used for image segmentation, particularly in medical imaging. It features a symmetric encoder-decoder structure with skip connections, allowing it to capture both spatial and contextual information effectively. The encoder compresses the input image into a low-dimensional representation, while the decoder gradually reconstructs the image, segmenting it into distinct regions. Skip connections between corresponding layers in the encoder and decoder help preserve detailed spatial information, leading to precise segmentation boundaries. U-Net's efficiency and accuracy make it popular in tasks like tumor detection, organ segmentation, and other applications requiring pixel-level classification.

2. Methodology

2.1. Classification

For classification swin transformer is used for classification on chestx-ray 14. The classification is done using ImageFolder. Inital pre-processing was done to use ImageFolder function. The images were normaslized and reshaped into size of 224 by 224. Batch size is 32, learning rate is 1e-5, number of classes is 13 and number of epochs is 15. The model was ran twice for with pretraining and without pretraining.

2.2. Localization

For localization swin transformer and tbx11-k dataset is used. The images are normalized and resized into size of 224 by 224. The model is trained for 10 epochs. Batch size is 2. Leanring rate is 1e-4. Initially by using swin transformer froc was not accurate. To improve the performance swin transformer is combined with faster cnn as faster cnn is better at predicting bounding boxes. AdamW optimizer is used. Swin transformer is used as backbone.

2.3. Segmentation

For segmentation U-Net is used on ChestX-Det dataset. Batch size is 4. Dice score is used for evaluation.

3. Results

3.1. Classification

For pretrained model the AUC is 0.6722.

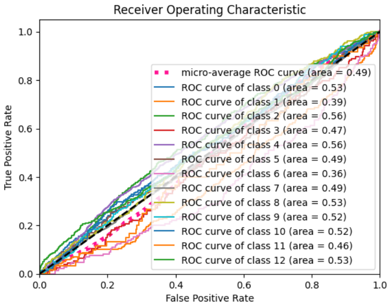

Figure 1. ROC curve for Pretrained Model

For model without pretraining AUC is 0.5522.

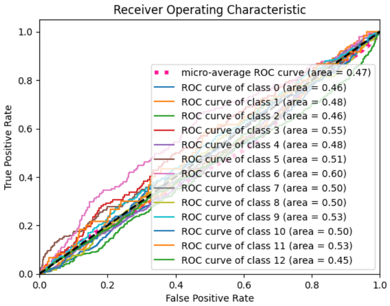

Figure 2. ROC curve for without Pretraining Model

3.2. Localization

FROC AUC is 0.655.

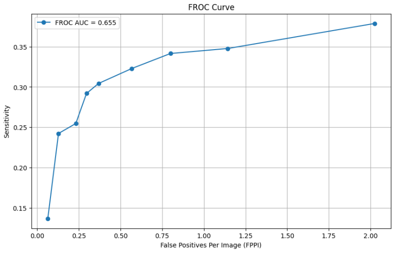

Figure 3. FROC curve

3.3. Segmentation

The model is still in training process.

4. References

1. https://arxiv.org/abs/2103.14030
2. https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community
3. https://arxiv.org/abs/2103.14030
4. https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community
5. Xie et al._2021_Self-Supervised Learning with Swin Transformers_Unknown-1.pdf
6. https://arxiv.org/abs/1506.01497
7. https://arxiv.org/abs/1505.04597
8. https://paperswithcode.com/dataset/chestx-det