Milestone Deep Learning

Group 09
Guilherme Cepeda 62931
Pedro Serrano 54853

6th of May 2024

Title : Transformer-Based Detection of B-Lines in Lung Ultrasound

1 Changes from project proposal

Before we start our introduction, we must explain that since the project proposal delivery our project objective changed. The previous title was *Transformer-Based Multi-Modal Data Fusion Method for COPD Classification*, it was not possible to continue with this idea as there wasn't any available multi-modal dataset. So to solve this dataset problem we had to change topics, and use the dataset regarding the paper that we will base our search on ***Deep Learning for Detection and Localization of B-Lines in Lung Ultrasound*** [1].

2 Introduction

Lung ultrasonography (LUS) has emerged as an important bedside imaging modality to support diagnostic assessment and therapeutic management in acute care settings. B-line artifacts in LUS are defined as hyperechoic lines that originate from the pleura and extend radially to the bottom of the screen, while moving synchronously with respiration. Despite the artifactual nature, B-lines play a key role in detecting and assessing the severity of pulmonary congestion in patients with disease states such as decompensated heart failure, chronic kidney disease on dialysis, viral and bacterial pneumonia, or interstitial lung disease. In clinical practice, physicians generally assign scores for disease severity based on a visual estimate of the quantity of B-lines.

In the paper [1], the authors studied several deep learning networks to attempt to classify, and count the number of B-lines by analysing a series of Lung Ultra-sounds of each patient. However, most of the models used were based around Convolutional Neural Networks (CNNs). Given the rising popularity of transformer-based architecture-based models, we will be attempting to recreate the results achieved in the paper, using such tools.

3 Data origin, pre-processing, and preparation

For this problem we will use the the Boston Emergency Department Lung UltraSound (BEDLUS) dataset, which is the same as the original paper [1]. The data was collected from patients admitted to the emergency department of Brigham and Women's Hospital, between November 2020, and March 2021 with symptoms related to flu-like illness. A total of 113 patients were examined and , typically, performed 12 Lung Ultra-sound exams, yielding a total of 1419 mp4 videos, with a total of 188670 frames. We should note that, due to patient discomfort, not everyone was subjected to the full 12 LUS, and for others, multiple shots of the same position were taken, giving us some inconsistencies in the data.

In this paper, it was introduced a novel way to identify B-lines, by their point of origin. Therefore, two LUS experts were tasked with identifying all B-lines throughout all videos of all patients, by indicating its point of origin. Since successive frames will most likely present the same B-lines, and in an effort to simplify the experts workload, only one in every four frames was analysed. A total of 15775 B-lines were annotated, on 10371 frames. Roughly, half of the clips (719) were annotated with the presence of B-lines, where the other half (700) were not.

In the paper [1], the authors performed state-of-the-art deep neural network architectures to attempt to predict the presence of B-lines, at three different levels - clip, frame and pixel. In the clip-level analysis, the model took a clip of 16 frames, to try and output a probability for the existence of B-lines in the LUS, where in the frame-level, the model input was each individual frame. The prediction for the entire video was simply the highest prediction of each frame. In the pixel level analysis, the main goal was a segmentation task, with the objective of localizing where the B-lines would originate. In the paper, the training test split was taken equally from the annotated, and non-annotated batches, meaning each split had an equal amount (except for odd splits, and the fact that there are slightly more positive cases than negative) of positive and negative cases.

4 Problem Statement and Technical Approach

The main focus of this project, will be to explore the transformer architecture, by attempting to recreate the results of this paper using such tools. Given the relatively small time scale, our primary goal is to do so, only on the detection of B-lines at the frame level.

Alongside the data provided, we also have the code for the data pre-processing, and preparation, which we will use. As for the specific model we will implement, we intend to test several common vision transformer architectures (ViT), like DeiT (Data-efficient Image Transformer) or the CaiT (Class attention image transformer), using frameworks like PyTorch and TensorFlow. These models are still not final, and may still be changed, as we still do not possess the sufficient knowledge to understand what models should be used. Details for the actual architecture can be seen below.

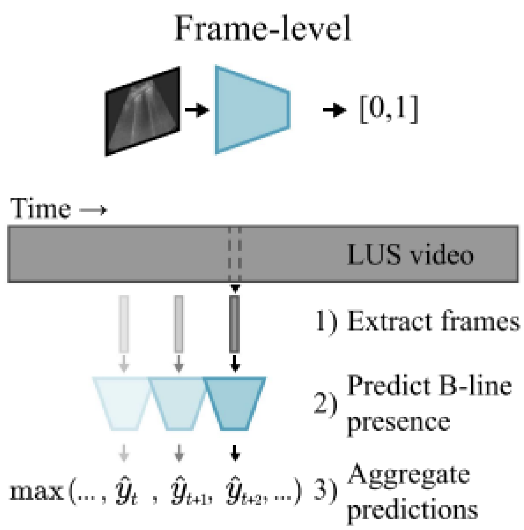

Figure 1: Frame Level Architecture

Figure 2: Frame Level Architecture: Vision Transformer

5 Experiments

Given the fact that our goal is simply to recreate the results of the paper using newer technologies, we intend to keep most of everything roughly the same. For the frame-level analysis, they used a lot of different encoders to find a probability of there being B-lines in an image. To compare the results, they calculated the F1 score, and the area under the curve (AUC). Given that each frame-level analysis outputs a probability, and the labels given were whether or not there were B-lines, to classify each clip as a true/false positve/negative, a certain threshold was given to the model, such that probabilities over the threshold would classify themselves as a positive prediction, and below, negative. Each network was run 5 different times, and the averages for both metrics were taken. For frame-level prediction, the best scores achieved in [1] were rounding the 0.922 for the AUC, using the EfficientNet-B0. This will be the result we will attempt to achieve.

References

[1] Ruben T. Lucassen, Mohammad H. Jafari, Nicole M. Duggan, Nick Jowkar, Alireza Mehrtash, Chanel Fischetti, Denie Bernier, Kira Prentice, Erik P. Duhaime, Mike Jin, Purang Abolmaesumi, Friso G. Heslinga, Mitko Veta, Maria A. Duran-Mendicuti, Sarah Frisken, Paul B. Shyn, Alexandra J. Golby, Edward Boyer, William M. Wells, Andrew J. Goldsmith, and Tina Kapur. Deep learning for detection and localization of b-lines in lung ultrasound. *IEEE Journal of Biomedical and Health Informatics*, 27(9):4352–4361, 2023.