Zero-Shot Object Detection for Chest X-Rays
Ellie Talius, Ruhi Sayana
Department of Computer Science, Stanford University

Introduction

- Using deep learning for object detection in medical imaging is challenging
  ○ Requires large amounts of labelled data
  ○ Expensive, time-consuming to annotate
- OpenAI’s CLIP model uses contrastive learning to build associations between images and text → can be used to understand medical image and radiology report pairs
  ○ Promising for zero-shot object detection
- No previous research on performing zero-shot object detection on chest x-ray images

Methods

- **Training**: Train CLIP architecture on a mini-batch of image-text pairs
  ○ Model learns to maximize cosine similarity between correct pairs and minimize cosine similarity of incorrect pairs
- **Evaluation**: Use learned image and text encoders to predict labelled bounding boxes
  ○ Use superpixel segmentation or selective search to generate RoI’s
  ○ Encode image crops with image encoder and labels with text encoder
  ○ Return boxes with similarity greater than threshold (t=0.5)

Datasets

- **Training**: MIMIC-CXR (300,000 images and reports)
- **Test**: Kaggle VinBigData Chest X-ray Abnormalities detection (3,000 images)
- Preprocessing: Resizing to 224 x 224, and 3 data augmentation set ups (random crop, random horizontal flip, full transformation pipeline)

Experiments

- **Pathology Detection**
  - **Zero-Shot vs. Fully Supervised**
    | Supervision | Model       | mAP Score |
    |-------------|-------------|-----------|
    | Fully Supervised | Detection 1 | 0.235     |
    | CLIP-Multiple-No Train | 0.037     |
    | CLIP-Multiple-No Train | 0.045     |
    | Best CLIP | 0.052     |

- **Text Query**
  - Prompts
    1) a chest x-ray with { } = 0.036
    2) part of a chest x-ray with { } = 0.025
    3) { } present in a chest x-ray = 0.035
    4) crop of a chest x-ray showing { } = 0.027

- **Data Augmentation**
  - Data Augmentation
    | mAP Score |
    |-----------|
    | Random Crop | 0.035     |
    | Random Flip | 0.025     |
    | All Transforms | 0.045     |

- **Region Proposal Selection**
  - Region Proposal Method
    | mAP Score |
    |-----------|
    | Superpixel Segmentation | 0.033     |
    | Selective Search | 0.042     |

Conclusions + Future Work

- We are able to perform zero-shot object detection for pathologies in chest x-ray images
- Key takeaway: zero-shot object detection can greatly benefit the medical field by helping automate and verify chest x-ray diagnosis without the need for expensive labelled data for training
- Our results are currently limited by the amount of compute (restricting batch size to a maximum of 128) and the use of a slow region proposal and object detection pipeline
- **Future steps**:
  ○ Developing a region proposal method that operates on the model image embeddings instead of raw pixels
  ○ Switching to a queue-based method for training to avoid the need for a large batch size

References