On the Importance of Image Encoding in Automated Chest X-Ray Report Generation

Otabek Nazarov, Mohammad Yaqub, Karthik Nandakumar ✉ otabe.k.nazarov@mbzuai.ac.ae

Abstract

• Compare different image encoding approaches (direct, fine-grained, CLIP, and Cluster-CLIP) along with multiple decoders to understand the relative importance of encoder and decoder components.

• Propose a novel cluster CLIP visual encoder (CCVE) that aims to generate more discriminative and explainable representations.

Introduction

• Clinical problem. Shortage of radiologists for on time chest X-ray diagnosis.

• Problem statement. Given an image of chest X-ray, generate a report capturing abnormalities.

• Existing works. Primarily focus on improving decoder and training method, but image encoding is neglected; mainly simply pretrained CNN is used.

• Dataset. MIMIC-CXR: ~200,000 image-report pairs.

Method

• Direct Visual Encoder (DVE). DenseNet-121 trained end-to-end along with decoder.

• Fine-Grained Visual Encoder (FVE). Two ConvNext-small classifiers (coarse with 14 classes and fine-grained with 410 classes).

• CLIP Visual Encoder (CVE). Contrastive language-image pretraining (CLIP) model trained on reports’ impression section.

• Cluster CLIP Visual Encoder (CCVE). Novel encoding method designed to produce distinct class embeddings. Image passed through convolutional filter prior to CLIP encoding; filters are selected based on image label during training stage; all filters are used during inference (see Figure 2).

• Decoders. Three different decoders are used: transformer, M2, and hierarchical RNN.

Results and Discussion

• FVE showed the best performance; thus, semantic information extraction is a key for effective image encoding.

• CLIP-based performed poorly
  • CCVE gave ROC-AUC of 0.71, while FVE gave ROC-AUC of 0.83
  • Contrastive training might be focusing on wrong words during training

• Future work.
  • CLIP-based methods need proper training method for medical data
  • Explore other methods that better extract semantic information.