IMPROVING AND OPTIMIZING MULTIMODAL MODELS FOR AUTOMATED X-RAY REPORT GENERATION

Minh Quan Tran[1]

[1] University of Information Technology. Vietnam National University, Ho Chi Minh City.

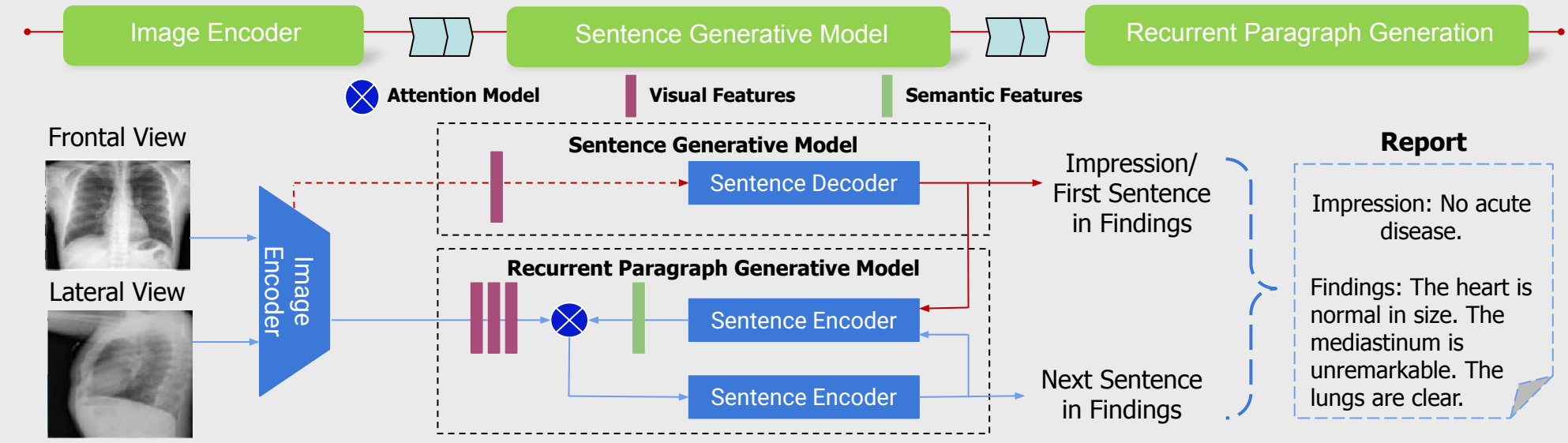
Motivations

Previous research mainly used deep learning models such as CNN and RNN to generate reports from X-ray images, but still had difficulty in combining multi-modal information and low accuracy. The goal of this project is to **improve these models** by applying *Vision Transformers* and *BioBERT*, **improving semantic understanding and image processing**. The project will s**olve the problem of overfitting** and **increase accuracy**, thereby supporting doctors in diagnosis and decision making.

Targets

- **Improve accuracy:** Optimize the ability to generate reports from X-ray images using multimodal models, improving diagnostic accuracy.
- **Minimize overfitting:** Use large and diverse data sets to improve the generalization ability of the model, ensuring reliability.
- **Support clinicians:** Develop automated systems to help doctors analyze images and make decisions more quickly and accurately.

Overview



Description

1. Image Encoder

- An image encoder is first applied to extract both global and regional visual features from the input images.
- The image encoder is a Convolutional Neural Network (CNN) that automatically extracts hierarchical visual features from images

2. Sentence Generative Model

- In general, both the one-sentence impression and the first sentence in the findings paragraph contain some high level descriptions of the image.
- A sentence generative model that takes the global visual features learned by the image encoder as input.
- Such a model can be trained to generate the impression. It can also be jointly trained with the recurrent generative model to generate the first sentence in the findings as an initialization of the recurrent model

3. Recurrent Paragraph Generation

- Recurrent paragraph generative model takes the sentence and regional image features as input and generates findings paragraph sentence by sentence.
- It has two main components: sentence encoder and attentional sentence decoder. Sentence Encoder is used to extract semantic vectors from text descriptions. Attentional Sentence Decoder takes regional visual features and the previously generated sentence as a multimodal input, and generates the next sentence.

4. Research Plan

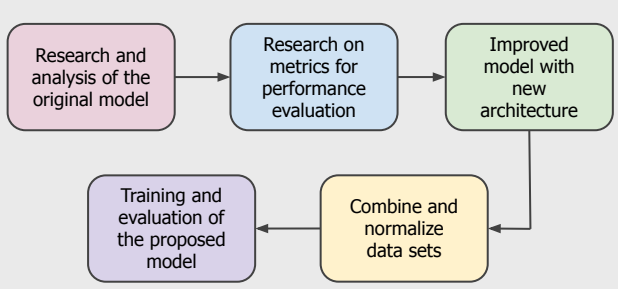


Figure 2. Research plan diagram

Recurrent Attention

Findings: The heart size and mediastinal contours appear within normal limits. No focal airspace consolidation, pleural effusion or pneumothorax. No acute bony abnormalities.
Impression: No acute cardiopulmonary finding.

Findings: The heart size and mediastinal silhouette are within normal limits for contour. The lungs are clear. No focal airspace consolidation. No pleural effusion or pneumothorax. Normal cardiomediastinal silhouette. Heart size is normal.
Impression: Clear lungs. No acute cardiopulmonary abnormality.

Ground Truth

Findings: The heart size and mediastinal silhouette are within normal limits for contour. The lungs are clear. No pneumothorax or pleural effusions. The XXXX are intact.
Impression: No acute cardiopulmonary abnormalities.

Note that, Findings is a paragraph containing some descriptive sentences; Impression is a conclusive sentence. XXXXs are wrongly removed keywords due to de-identification.

Findings: Mediastinal contours are within normal limits. Heart size is within normal limits. No focal consolidation, pneumothorax or pleural effusion. No bony abnormality. Vague density in right mid lung, XXXX related to scapular tip and superimposed ribs. Not visualized on lateral exam.
Impression: Vague density in right XXXX, XXXX related to scapular tip and superimposed ribs. Consider oblique images to exclude true nodule

Figure 1. Example of original report compared to report generated by recurrent attention model

NII

Minh Quan Tran – University of Information Technology. Vietnam National University. Ho Chi Minh City.
TEL : 0337459509 Email : 22521191@gm.uit.edu.vn