Hướng phát triển :

Dựa theo bài báo <https://arxiv.org/pdf/2201.12944> sẽ có các hướng sau để phét triển model image captioning :

* Attention-Based method (Chủ yếu dựa trên CNN và RCNN kết hợp với LSTM )

A diagram of a machine

Description automatically generated

Kiến trúc của Attention-Based Method

* Link code và bài báo
* AoANet : <https://github.com/husthuaan/AoANet>
* Link bài báo : <https://arxiv.org/abs/1908.06954>

Graph-Based method for Spatial and Sematic Relation between Image Elements

(CNN + GCN + LSTM)

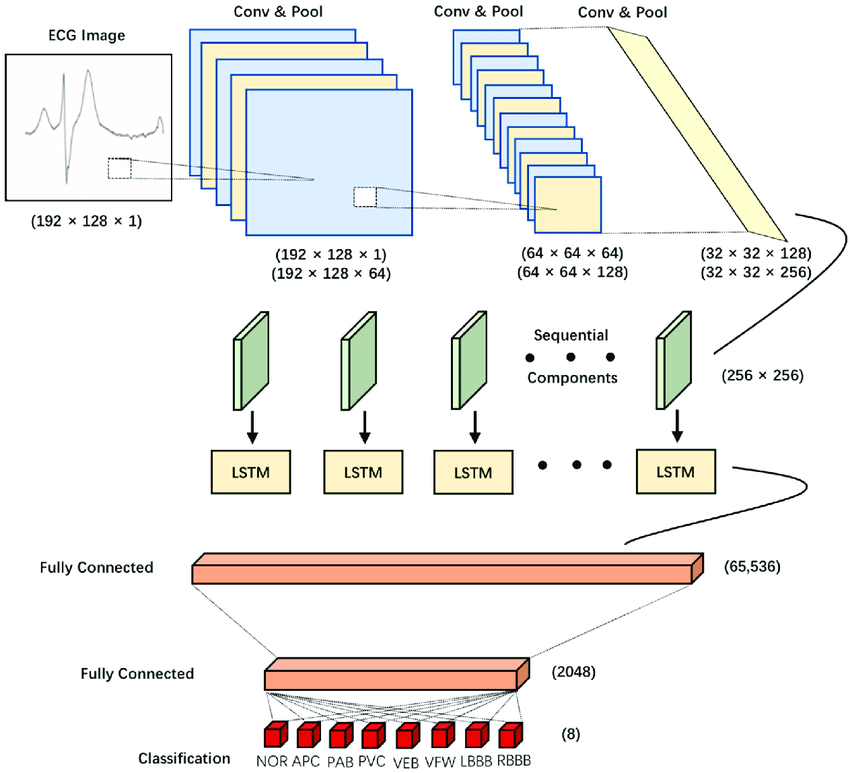
A diagram of a graph

Description automatically generated

Kiến trúc graph-based method

Convolution Network-Based method (CNN method )

Bài báo so sánh các CNN model : <https://arxiv.org/pdf/2102.11506>



Link bài báo về (MobileNetV3) :[1905.02244 (arxiv.org)](https://arxiv.org/pdf/1905.02244)

A diagram of a block diagram

Description automatically generated

A diagram of a block diagram

Description automatically generated

* Xử lý ảnh trong model:
* Ảnh đầu vào sẽ thường được điều chỉnh kích thước và chuẩn hóa trước khi đưa vào model.
* MobileNetV3 sử dụng các convolutional layers để trích xuất các đặc trưng từ ảnh đầu vào
* Sau khi ảnh đi qua các lớp tích chập ( convolutional layers) và các lớp kết nối (fully connected layers), model sẽ tạo ra một vector đặc trưng đại diện cho các thông tin quan trọng trong ảnh như hình dạng, kết cấu, màu sắc, đối tượng.
* Tạo văn bản: Sau khi có vector đặc trưng thì có thể kết hợp với một số model decoder như LSTM hoặc Transformer để tạo ra từng từ một cho đến khi hoàn thành câu mô tả ảnh

A diagram of a diagram

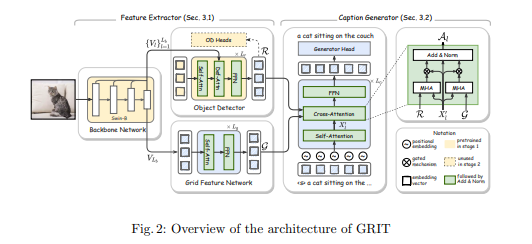
Description automatically generated

Link bài báo (baby talk ) RNN + LSTM : <https://arxiv.org/pdf/1803.09845>

Link bài báo NIC ( CNN + Target Caption Model ) : <https://arxiv.org/pdf/1712.02051>

Transformer-Based method   
- Có thể sử dụng GRIT model do sử dụng 1 V100 GPU

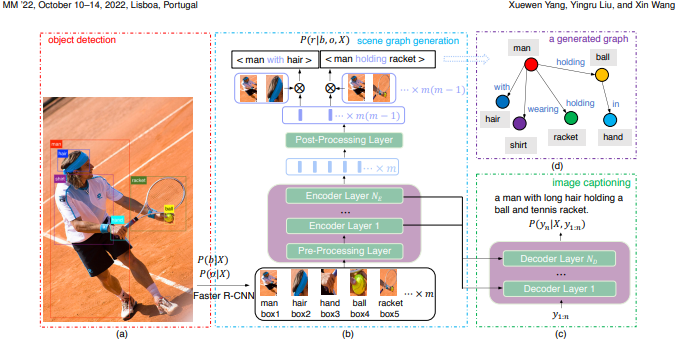
Link bài báo <https://arxiv.org/pdf/2207.09666>



Combining Transformers and Scene Graphs

Link bài báo ( Reformer model ) : <https://arxiv.org/pdf/2107.14178>

Note ( chưa chắc về tính khả thi do thiếu thông tin về tài nguyên training )



Các hướng khác ngoài bài báo :

EfficientNet : Link bài báo : <https://www.researchgate.net/publication/366146850_Image_Caption_Generator_using_EfficientNet>

BUTD (bottom up top down) : <https://github.com/IBM/Image-Captioning-Attack>

Link bài báo :<https://arxiv.org/pdf/1707.07998>

A diagram of a software algorithm

Description automatically generated

Kiến trúc BUTD

AoANet : <https://github.com/husthuaan/AoANet>

Link bài báo : <https://arxiv.org/abs/1908.06954>

A screenshot of a computer

Description automatically generated

A diagram of a software algorithm

Description automatically generated

Show and Tell : <https://arxiv.org/pdf/1411.4555v2>

A diagram of a machine

Description automatically generated

NIC ( CNN + Target Caption Model ) :

* Link bài báo : <https://arxiv.org/pdf/1712.02051>