M2 : Meshed-Memory Transformer for Image Captioning

Nguyễn Minh Châu

University of Information Technology
HCMC, Vietnam

What ?

We introduce an image captioning algorithm, in which we have:

- Proposed a novel image captioning algorithm to generating caption from the objects in the image based on the **robust** Transformer architecture.
- Compare our model with different fully-attentive architectures for image captioning.

Why ?

- **Image captioning** can assist **social media** platforms with determining the content of images uploaded by users. This act may be useful in resolving issues such as **incorrect information**, **sensual behavior, and unsuitable standards**. Image captioning may also assist people **pick the best image by inputting keywords**.
- Transformer architecture is used in most of state-of-the-art methods in both Computer Vision and Natural Language Processing

Overview

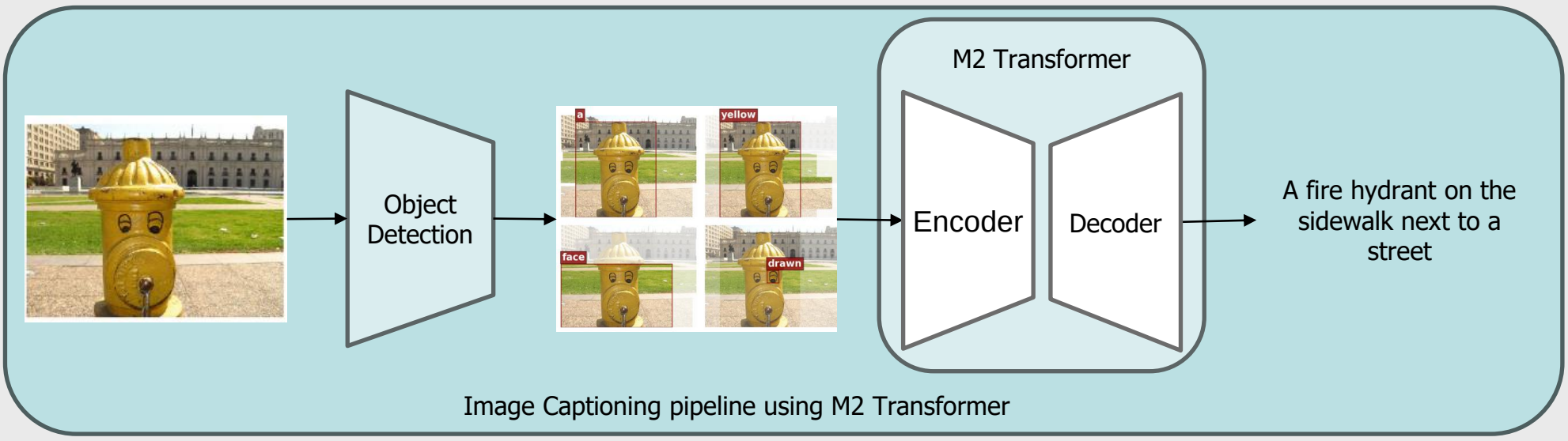

Image Captioning pipeline using M2 Transformer

Description

1. Image Captioning

- Image Captioning is a problem in which the model has to generate the caption that related to the input image.
- Image Captioning is a hard problem due to the combination of two big tasks: Computer Vision and Natural Language Processing.

Samples from COCO Captions (Image Captioning Dataset)

2. Object Detection

- The objects in the input image can be detected using the object detection model.
- The output of the object detection model are several detected objects. Those detected objects will be put to the Captioning Generation module.

3. Captioning Generation

- We used Transformer architecture as a base to generate the caption.
- The detected objects are the inputs for the M2 Transformer.
- The output of the M2 Transformer will be the sequence of words that describes the content of the input image

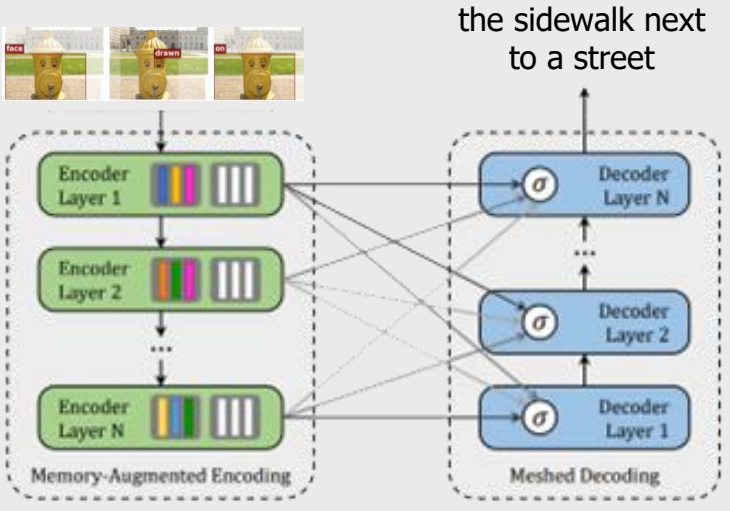

Image Captioning using Transformer architecture

4. Expected Results

- A report about the M2-Transformer, the experiment result in which comparing the the M2-Transformer with another variants.
- A Graphical User Interface for easier usage.

NII

Nguyễn Minh Châu– Trường Đại học Công nghệ Thông tin
TEL : 0326342254 Email : chaunm.17@uit.edu.vn